GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 11:36:26 ; Search time 4832.99 Seconds
(without alignments)
11760.106 Million cell updates/sec

Title: US-09-882-986-1
Perfect score: 2716
Sequence: 1 aaagagacgaactgaagaaa.....ctttaaaaaaaaaaaaaaaa 2716

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2716	100.0	2716	6	AX341991	AX341991 Sequence
2	2716	100.0	2716	8	AF272891	AF272891 Arabidops
3	2700	99.4	2834	8	AF268377	AF268377 Arabidops
4	2697.4	99.3	2735	8	AY064960	AY064960 Arabidops
5	990.2	36.5	82403	8	AC007063	AC007063 Arabidops
6	978.2	36.0	3009	8	AY017415	AY017415 Oryza sat
7	69.2	2.5	65076	8	NCB13O20	AL451015 Neurospor
8	65	2.4	3228	3	DME238970	AJ238970 Drosophil
9	61.8	2.3	2975	3	AY051796	AY051796 Drosophil
10	57.2	2.1	3066	9	HUMNCBP	D32002 Human mRNA
11	56.8	2.1	14831	2	AC013864	AC013864 Drosophil
12	56.8	2.1	160622	3	AC023715	AC023715 Drosophil
c 13	56.8	2.1	167089	3	AC023747	AC023747 Drosophil
14	56.8	2.1	311464	3	AE003431	AE003431 Drosophil
15	56.2	2.1	163029	9	AC022016	AC022016 Homo sapi
16	55.6	2.0	2541	9	BC001450	BC001450 Homo sapi
17	55.6	2.0	2828	9	HSCBP80	X80030 H.sapiens C
18	53.2	2.0	28440	8	SPAC6G10	Z98603 S.pombe chr
19	50.8	1.9	162770	3	AC008309	AC008309 Drosophil
20	50.8	1.9	167475	3	AC016132	AC016132 Drosophil
c 21	50.8	1.9	220936	2	AC020256	AC020256 Drosophil
22	50.8	1.9	240910	3	AE003734	AE003734 Drosophil
c 23	48.4	1.8	78396	9	AL590486	AL590486 Human DNA
24	46	1.7	125020	9	AF429315	AF429315 Homo sapi
c 25	46	1.7	293431	2	PFMAL13P4	AL049181 Plasmodiu
26	45.6	1.7	372	3	AF162337	AF162337 Mytilus g
c 27	45.4	1.7	603	8	AF464753	AF464753 Pholidota
c 28	45.4	1.7	6242	6	AX347050	AX347050 Sequence
29	45.4	1.7	74371	9	AC005369	AC005369 Homo sapi
c 30	45.2	1.7	7218	6	I66494	I66494 Sequence 14
31	45	1.7	1386	6	AR075041	AR075041 Sequence
32	45	1.7	1386	6	AR141859	AR141859 Sequence
33	45	1.7	2175	10	BC007144	BC007144 Mus muscu
c 34	45	1.7	3780	3	AF238312	AF238312 Dictyoste
35	44.8	1.6	125021	8	AC007504	AC007504 Arabidops
36	44.6	1.6	9330	3	AF466305	AF466305 Dictyoste
37	44.4	1.6	2062	3	AF035482	AF035482 Trichoplu
c 38	44.4	1.6	6552	6	AX251432	AX251432 Sequence
39	44.2	1.6	1761	10	BC022641	BC022641 Mus muscu
c 40	44.2	1.6	6063	6	AX323780	AX323780 Sequence
c 41	44.2	1.6	99053	8	AC005292	AC005292 Genomic s
c 42	44.2	1.6	161619	9	AC046174	AC046174 Homo sapi
c 43	44.2	1.6	167106	2	AC090102	AC090102 Homo sapi
44	44.2	1.6	293431	2	PFMAL13P4	AL049181 Plasmodiu
c 45	44.2	1.6	349980	6	AX344572	AX344572 Sequence

ALIGNMENTS

RESULT 1
AX341991
LOCUS AX341991 2716 bp DNA linear PAT 10-JAN-2002
DEFINITION Sequence 1 from Patent WO0196585.
ACCESSION AX341991
VERSION AX341991.1 GI:18137972
KEYWORDS .
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (sites)
AUTHORS Schroeder,J., Hugouvieux,V. and Kwak,J.M.
TITLE Modulation of abscisic acid signal transduction in plants
JOURNAL Patent: WO 0196585-A 1 20-DEC-2001;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA (US)
FEATURES Location/Qualifiers
source 1. .2716
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/note="abscisic acid (ABA) hypersensitive (ABH1) cDNA"
CDS 37. .2583

```
                     /note="ABH1 protein"
                     /codon_start=1
                     /protein_id="CAD20687.1"
                     /db_xref="GI:18137973"
                     /translation="MSNWKTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGD
                     QVLPFLLQCAEQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSI
                     RILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILS
                     SLPWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVE
                     DLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYS
                     GKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASY
                     MANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAV
                     RALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEIL
                     QREVRLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKV
                     KEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSK
                     LCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFH
                     VSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLV
                     EGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLFQS
                     FLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLS
                     TLGYLTAFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAISSALQFPLH"
BASE COUNT      778 a    532 c    628 g    778 t
ORIGIN
```

Query Match 100.0%; Score 2716; DB 6; Length 2716;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2716; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 aaagagacgaactgaagaaaaacctctcggaagaagatgagcaattggaaaactcttctc 60
Db       1 AAAGAGACGAACTGAAGAAAAACCTCTCGGAAGAAGATGAGCAATTGGAAAACTCTTCTC 60

Qy      61 cttcgcatcggcgaaaagggacctgagtacggcacttcctccgactacaaagaccacatc 120
Db      61 CTTCGCATCGGCGAAAAGGGACCTGAGTACGGCACTTCCTCCGACTACAAAGACCACATC 120

Qy     121 gagacttgtttcggtgtcattcgtagagaaatcgagcgttctggagatcaagttttgcct 180
Db     121 GAGACTTGTTTCGGTGTCATTCGTAGAGAAATCGAGCGTTCTGGAGATCAAGTTTTGCCT 180

Qy     181 tttctactacaatgtgctgaacaattgcctcataagattcctttgtatgggactttgatt 240
Db     181 TTTCTACTACAATGTGCTGAACAATTGCCTCATAAGATTCCTTTGTATGGGACTTTGATT 240

Qy     241 ggtttgttgaacttggagaatgaagattttgtccagaagctagtagaaagtgtccacgct 300
Db     241 GGTTTGTTGAACTTGGAGAATGAAGATTTTGTCCAGAAGCTAGTAGAAAGTGTCCACGCT 300

Qy     301 aatttccaggtcgctttagattctggcaactgcaacagtatccgtatattgcttcgcttt 360
Db     301 AATTTCCAGGTCGCTTTAGATTCTGGCAACTGCAACAGTATCCGTATATTGCTTCGCTTT 360

Qy     361 atgacttccctgttgtgcagtaaggttattcaacctgcttctttgattgtcgtcttcgaa 420
Db     361 ATGACTTCCCTGTTGTGCAGTAAGGTTATTCAACCTGCTTCTTTGATTGTCGTCTTCGAA 420

Qy     421 acattgctatcatctgctgccactactgtggatgaagagaaaggaaatccatcatggcag 480
Db     421 ACATTGCTATCATCTGCTGCCACTACTGTGGATGAAGAGAAAGGAAATCCATCATGGCAG 480

Qy     481 ccacaagctgacttttacgttatatgcatcttgtccagcctcccgtggggaggatcagaa 540
Db     481 CCACAAGCTGACTTTTACGTTATATGCATCTTGTCCAGCCTCCCGTGGGGAGGATCAGAA 540

Qy     541 ctcgctgagcaagttcctgatgagattgaaagagtgttagttgggatacaagcttatttg 600
Db     541 CTCGCTGAGCAAGTTCCTGATGAGATTGAAAGAGTGTTAGTTGGGATACAAGCTTATTTG 600

Qy     601 agcatccgaaagaattcttccacctctgggttaaacttttttcacaacggagaatttgaa 660
Db     601 AGCATCCGAAAGAATTCTTCCACCTCTGGGTTAAACTTTTTTCACAACGGAGAATTTGAA 660

Qy     661 agcagccttgcagagaaggatttcgtggaggatctattggatcgaattcagtctctggct 720
```

```
Db     661 AGCAGCCTTGCAGAGAAGGATTTCGTGGAGGATCTATTGGATCGAATTCAGTCTCTGGCT 720

Qy     721 tccaatggatggaaacttgaaagcgtacctaggcctcatctctcgtttgaagctcaactc 780
Db     721 TCCAATGGATGGAAACTTGAAAGCGTACCTAGGCCTCATCTCTCGTTTGAAGCTCAACTC 780

Qy     781 gttgctggaaagtttcatgagctacgtcccattaaatgtatggaacaaccgagtccacct 840
Db     781 GTTGCTGGAAAGTTTCATGAGCTACGTCCCATTAAATGTATGGAACAACCGAGTCCACCT 840

Qy     841 tctgatcattcgagggcatacagtggcaagcaaaagcatgatgcattgacgagatatccc 900
Db     841 TCTGATCATTCGAGGGCATACAGTGGCAAGCAAAAGCATGATGCATTGACGAGATATCCC 900

Qy     901 cagagaattcgtaggttgaatatatttccagctaataaaatggaggatgtacaaccaatt 960
Db     901 CAGAGAATTCGTAGGTTGAATATATTTCCAGCTAATAAAATGGAGGATGTACAACCAATT 960

Qy     961 gatcgttttgtcgtggaggagtatttgctggatgtgctcttctatttgaatggatgtcgg 1020
Db     961 GATCGTTTTGTCGTGGAGGAGTATTTGCTGGATGTGCTCTTCTATTTGAATGGATGTCGG 1020

Qy    1021 aaggagtgtgcatcctacatggctaatcttcctgttacatttcggtacgagtatcttatg 1080
Db    1021 AAGGAGTGTGCATCCTACATGGCTAATCTTCCTGTTACATTTCGGTACGAGTATCTTATG 1080

Qy    1081 gcagagacactattttctcagatactgctgctaccccagccaccattcaagactctttat 1140
Db    1081 GCAGAGACACTATTTTCTCAGATACTGCTGCTACCCCAGCCACCATTCAAGACTCTTTAT 1140

Qy    1141 tatacactcgtgattatggatctttgtaaggctcttccgggtgccttttcctgctgttgtt 1200
Db    1141 TATACACTCGTGATTATGGATCTTTGTAAGGCTCTTCCGGGTGCCTTTCCTGCTGTTGTT 1200

Qy    1201 gctggcgctgttcgtgcactatttgagaaaatatccgacttagacatggaatccaggacg 1260
Db    1201 GCTGGCGCTGTTCGTGCACTATTTGAGAAAATATCCGACTTAGACATGGAATCCAGGACG 1260

Qy    1261 cgtcttatcctctggttttctcaccacttatccaacttccaattcatctggccgtgggaa 1320
Db    1261 CGTCTTATCCTCTGGTTTTCTCACCACTTATCCAACTTCCAATTCATCTGGCCGTGGGAA 1320

Qy    1321 gagtgggcttttgtgttggatcttcccaagtgggcccctaagcgtgtatttgttcaggag 1380
Db    1321 GAGTGGGCTTTTGTGTTGGATCTTCCCAAGTGGGCCCCTAAGCGTGTATTTGTTCAGGAG 1380

Qy    1381 attttgcaaagagaagtacgcttgtcttactgggataaaattaagcagagcattgagaat 1440
Db    1381 ATTTTGCAAAGAGAAGTACGCTTGTCTTACTGGGATAAAATTAAGCAGAGCATTGAGAAT 1440

Qy    1441 gcgactgccctagaagaattacttcctccaaaagctggtccgaattttatgtattccttg 1500
Db    1441 GCGACTGCCCTAGAAGAATTACTTCCTCCAAAAGCTGGTCCGAATTTTATGTATTCCTTG 1500

Qy    1501 gaagaaggtaaagagaaaacagaagaacagcaattgtcagccgaattgagcaggaaggtc 1560
Db    1501 GAAGAAGGTAAAGAGAAAACAGAAGAACAGCAATTGTCAGCCGAATTGAGCAGGAAGGTC 1560

Qy    1561 aaggaaaaacaaaccgcacgtgacatgatagtgtggattgaagaaacgatatatccagtt 1620
Db    1561 AAGGAAAAACAAACCGCACGTGACATGATAGTGTGGATTGAAGAAACGATATATCCAGTT 1620

Qy    1621 catggttttgaagttactcttacaatagttgtacagaccttacttgacatcggatcaaaa 1680
Db    1621 CATGGTTTTGAAGTTACTCTTACAATAGTTGTACAGACCTTACTTGACATCGGATCAAAA 1680

Qy    1681 agtttcactcatttggtcactgtcctggagcgatatggccaagtattttcaaagctttgt 1740
Db    1681 AGTTTCACTCATTTGGTCACTGTCCTGGAGCGATATGGCCAAGTATTTTCAAAGCTTTGT 1740

Qy    1741 cctgataacgataagcaggtgatgctattatctcaagtgagtacatactggaaaaacaat 1800
Db    1741 CCTGATAACGATAAGCAGGTGATGCTATTATCTCAAGTGAGTACATACTGGAAAAACAAT 1800
```

```
Qy   1801 gtacaaatgacggcggtggcaattgataggatgatgggttatagactagtatctaatcag 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 GTACAAATGACGGCGGTGGCAATTGATAGGATGATGGGTTATAGACTAGTATCTAATCAG 1860

Qy   1861 gcaattgttagatgggtgttctctccagaaaatgttgatcagtttcatgtgtctgatcag 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GCAATTGTTAGATGGGTGTTCTCTCCAGAAAATGTTGATCAGTTTCATGTGTCTGATCAG 1920

Qy   1921 ccatgggagatacttggcaatgctcttaacaagacttataaccgtatctctgatttgagg 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CCATGGGAGATACTTGGCAATGCTCTTAACAAGACTTATAACCGTATCTCTGATTTGAGG 1980

Qy   1981 aaagatatatcaaacattacgaaaaatgttttggttgctgagaaagcttcagccaatgca 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AAAGATATATCAAACATTACGAAAAATGTTTTGGTTGCTGAGAAAGCTTCAGCCAATGCA 2040

Qy   2041 cgagtagagttggaggctgctgagagcaaactttccctagtggaaggtgaacccgttctt 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CGAGTAGAGTTGGAGGCTGCTGAGAGCAAACTTTCCCTAGTGGAAGGTGAACCCGTTCTT 2100

Qy   2101 ggtgagaatccagcgaagatgaagcgtttaaaatcaacagtggagaagacaggggaagcg 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GGTGAGAATCCAGCGAAGATGAAGCGTTTAAAATCAACAGTGGAGAAGACAGGGGAAGCG 2160

Qy   2161 gagttatctcttcgggagtccctagaggcaaaagaggctcttcttaacagagctctctct 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 GAGTTATCTCTTCGGGAGTCCCTAGAGGCAAAAGAGGCTCTTCTTAACAGAGCTCTCTCT 2220

Qy   2221 gagaccgaggttttactgctcttgctgttccaaagtttcttaggtgtactgaaggaacgg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GAGACCGAGGTTTTACTGCTCTTGCTGTTCCAAAGTTTCTTAGGTGTACTGAAGGAACGG 2280

Qy   2281 ctcccagatccaactaaagtgagatcagtgcaggatctaaaatctataggtgctgaagat 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CTCCCAGATCCAACTAAAGTGAGATCAGTGCAGGATCTAAAATCTATAGGTGCTGAAGAT 2340

Qy   2341 gacaagccatctgcgatggacgtggacagcgagaatggaaacccaaagaagagttgcgaa 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GACAAGCCATCTGCGATGGACGTGGACAGCGAGAATGGAAACCCAAAGAAGAGTTGCGAA 2400

Qy   2401 gtcggtgagagagaacagtggtgcttatcaacacttggctatctcacggcatttacaagg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GTCGGTGAGAGAGAACAGTGGTGCTTATCAACACTTGGCTATCTCACGGCATTTACAAGG 2460

Qy   2461 caatatgcgagcgagatatggcctcacatggagaagttggagtcagaagtgttctcgggt 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 CAATATGCGAGCGAGATATGGCCTCACATGGAGAAGTTGGAGTCAGAAGTGTTCTCGGGT 2520

Qy   2521 gaagatgtgcatcctctctttctccaagccatatcttctgcacttcaattcccattacat 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GAAGATGTGCATCCTCTCTTTCTCCAAGCCATATCTTCTGCACTTCAATTCCCATTACAT 2580

Qy   2581 taatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgttatgagattct 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TAATCTTCCTCTTTCAATCTCAATCAAACCTGTCTCTTTTGTTTTTTGTTATGAGATTCT 2640

Qy   2641 gattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaagagtttaaacttt 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 GATTCTGACATCAAGTTATTAGGAAATTGAAAAGAGTCAAAAAACAAGAGTTTAAACTTT 2700

Qy   2701 aaaaaaaaaaaaaaaa 2716
          ||||||||||||||||
Db   2701 AAAAAAAAAAAAAAAA 2716
```

```
RESULT   2
AF272891
LOCUS       AF272891               2716 bp    mRNA    linear   PLN 06-SEP-2001
DEFINITION  Arabidopsis thaliana mRNA cap binding protein mRNA, complete cds.
ACCESSION   AF272891
VERSION     AF272891.1  GI:15192737
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2716)
  AUTHORS   Hugouvieux,V., Kwak,J.M. and Schroeder,J.I.
  TITLE     An mRNA Cap Binding Protein, ABH1, Modulates Early Abscisic Acid
            Signal Transduction in Arabidopsis
  JOURNAL   Cell 106 (4), 477-487 (2001)
  MEDLINE   21417142
   PUBMED   11525733
REFERENCE   2  (bases 1 to 2716)
  AUTHORS   Hugouvieux,V.G. and Schroeder,J.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-2000) Biology, UC San Diego, 9500 Gilman Drive,
            La Jolla, CA 92093-0116, USA
FEATURES             Location/Qualifiers
     source          1. .2716
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
     CDS             37. .2583
                     /note="similar to nuclear cap binding complex subunit GCR3
                     from Saccharomyces cerevisiae and to CBP80 from Hela
                     cells; large subunit of the cap binding complex"
                     /codon_start=1
                     /product="mRNA cap binding protein"
                     /protein_id="AAK91588.1"
                     /db_xref="GI:15192738"
                     /translation="MSNWKTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGD
                     QVLPFLLQCAEQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSI
                     RILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILS
                     SLPWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVE
                     DLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYS
                     GKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASY
                     MANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAV
                     RALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEIL
                     QREVRLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKV
                     KEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSK
                     LCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFH
                     VSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLV
                     EGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLFQS
                     FLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLS
                     TLGYLTAFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAISSALQFPLH"
BASE COUNT      778 a    532 c    628 g    778 t
ORIGIN
```

```
  Query Match           100.0%;  Score 2716;  DB 8;  Length 2716;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2716;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 aaagagacgaactgaagaaaaacctctcggaagaagatgagcaattggaaaactcttctc 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AAAGAGACGAACTGAAGAAAAACCTCTCGGAAGAAGATGAGCAATTGGAAAACTCTTCTC 60

Qy      61 cttcgcatcggcgaaaagggacctgagtacggcacttcctccgactacaaagaccacatc 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTTCGCATCGGCGAAAAGGGACCTGAGTACGGCACTTCCTCCGACTACAAAGACCACATC 120

Qy     121 gagacttgtttcggtgtcattcgtagagaaatcgagcgttctggagatcaagttttgcct 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GAGACTTGTTTCGGTGTCATTCGTAGAGAAATCGAGCGTTCTGGAGATCAAGTTTTGCCT 180

Qy     181 tttctactacaatgtgctgaacaattgcctcataagattcctttgtatgggactttgatt 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTTCTACTACAATGTGCTGAACAATTGCCTCATAAGATTCCTTTGTATGGGACTTTGATT 240

Qy     241 ggtttgttgaacttggagaatgaagattttgtccagaagctagtagaaagtgtccacgct 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGTTTGTTGAACTTGGAGAATGAAGATTTTGTCCAGAAGCTAGTAGAAAGTGTCCACGCT 300
```

```
Qy    301 aatttccaggtcgctttagattctggcaactgcaacagtatccgtatattgcttcgcttt 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AATTTCCAGGTCGCTTTAGATTCTGGCAACTGCAACAGTATCCGTATATTGCTTCGCTTT 360

Qy    361 atgacttccctgttgtgcagtaaggttattcaacctgcttctttgattgtcgtcttcgaa 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATGACTTCCCTGTTGTGCAGTAAGGTTATTCAACCTGCTTCTTTGATTGTCGTCTTCGAA 420

Qy    421 acattgctatcatctgctgccactactgtggatgaagagaaaggaaatccatcatggcag 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ACATTGCTATCATCTGCTGCCACTACTGTGGATGAAGAGAAAGGAAATCCATCATGGCAG 480

Qy    481 ccacaagctgacttttacgttatatgcatcttgtccagcctcccgtggggaggatcagaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCACAAGCTGACTTTTACGTTATATGCATCTTGTCCAGCCTCCCGTGGGGAGGATCAGAA 540

Qy    541 ctcgctgagcaagttcctgatgagattgaaagagtgttagttgggatacaagcttatttg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTCGCTGAGCAAGTTCCTGATGAGATTGAAAGAGTGTTAGTTGGGATACAAGCTTATTTG 600

Qy    601 agcatccgaaagaattcttccacctctgggttaaacttttttcacaacggagaatttgaa 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AGCATCCGAAAGAATTCTTCCACCTCTGGGTTAAACTTTTTTCACAACGGAGAATTTGAA 660

Qy    661 agcagccttgcagagaaggatttcgtggaggatctattggatcgaattcagtctctggct 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AGCAGCCTTGCAGAGAAGGATTTCGTGGAGGATCTATTGGATCGAATTCAGTCTCTGGCT 720

Qy    721 tccaatggatggaaacttgaaagcgtacctaggcctcatctctcgtttgaagctcaactc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TCCAATGGATGGAAACTTGAAAGCGTACCTAGGCCTCATCTCTCGTTTGAAGCTCAACTC 780

Qy    781 gttgctggaaagtttcatgagctacgtcccattaaatgtatggaacaaccgagtccacct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTTGCTGGAAAGTTTCATGAGCTACGTCCCATTAAATGTATGGAACAACCGAGTCCACCT 840

Qy    841 tctgatcattcgagggcatacagtggcaagcaaaagcatgatgcattgacgagatatccc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TCTGATCATTCGAGGGCATACAGTGGCAAGCAAAAGCATGATGCATTGACGAGATATCCC 900

Qy    901 cagagaattcgtaggttgaatatatttccagctaataaaatggaggatgtacaaccaatt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CAGAGAATTCGTAGGTTGAATATATTTCCAGCTAATAAAATGGAGGATGTACAACCAATT 960

Qy    961 gatcgttttgtcgtggaggagtatttgctggatgtgctcttctatttgaatggatgtcgg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GATCGTTTTGTCGTGGAGGAGTATTTGCTGGATGTGCTCTTCTATTTGAATGGATGTCGG 1020

Qy   1021 aaggagtgtgcatcctacatggctaatcttcctgttacatttcggtacgagtatcttatg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AAGGAGTGTGCATCCTACATGGCTAATCTTCCTGTTACATTTCGGTACGAGTATCTTATG 1080

Qy   1081 gcagagacactattttctcagatactgctgctaccccagccaccattcaagactctttat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GCAGAGACACTATTTTCTCAGATACTGCTGCTACCCCAGCCACCATTCAAGACTCTTTAT 1140

Qy   1141 tatacactcgtgattatggatctttgtaaggctcttccgggtgcctttcctgctgttgtt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TATACACTCGTGATTATGGATCTTTGTAAGGCTCTTCCGGGTGCCTTTCCTGCTGTTGTT 1200

Qy   1201 gctggcgctgttcgtgcactatttgagaaaatatccgacttagacatggaatccaggacg 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GCTGGCGCTGTTCGTGCACTATTTGAGAAAATATCCGACTTAGACATGGAATCCAGGACG 1260

Qy   1261 cgtcttatcctctggttttctcaccacttatccaacttccaattcatctggccgtgggaa 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CGTCTTATCCTCTGGTTTTCTCACCACTTATCCAACTTCCAATTCATCTGGCCGTGGGAA 1320

Qy   1321 gagtgggcttttgtgttggatcttcccaagtgggcccctaagcgtgtatttgttcaggag 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GAGTGGGCTTTTGTGTTGGATCTTCCCAAGTGGGCCCCTAAGCGTGTATTTGTTCAGGAG 1380

Qy   1381 attttgcaaagagaagtacgcttgtcttactgggataaaattaagcagagcattgagaat 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ATTTTGCAAAGAGAAGTACGCTTGTCTTACTGGGATAAAATTAAGCAGAGCATTGAGAAT 1440

Qy   1441 gcgactgccctagaagaattacttcctccaaaagctggtccgaattttatgtattccttg 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GCGACTGCCCTAGAAGAATTACTTCCTCCAAAAGCTGGTCCGAATTTTATGTATTCCTTG 1500

Qy   1501 gaagaaggtaaagagaaaacagaagaacagcaattgtcagccgaattgagcaggaaggtc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GAAGAAGGTAAAGAGAAAACAGAAGAACAGCAATTGTCAGCCGAATTGAGCAGGAAGGTC 1560

Qy   1561 aaggaaaaacaaaccgcacgtgacatgatagtgtggattgaagaaacgatatatccagtt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AAGGAAAAACAAACCGCACGTGACATGATAGTGTGGATTGAAGAAACGATATATCCAGTT 1620

Qy   1621 catggttttgaagttactcttacaatagttgtacagaccttacttgacatcggatcaaaa 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CATGGTTTTGAAGTTACTCTTACAATAGTTGTACAGACCTTACTTGACATCGGATCAAAA 1680

Qy   1681 agtttcactcatttggtcactgtcctggagcgatatggccaagtattttcaaagctttgt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 AGTTTCACTCATTTGGTCACTGTCCTGGAGCGATATGGCCAAGTATTTTCAAAGCTTTGT 1740

Qy   1741 cctgataacgataagcaggtgatgctattatctcaagtgagtacatactggaaaaacaat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CCTGATAACGATAAGCAGGTGATGCTATTATCTCAAGTGAGTACATACTGGAAAAACAAT 1800

Qy   1801 gtacaaatgacggcggtggcaattgataggatgatgggttatagactagtatctaatcag 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 GTACAAATGACGGCGGTGGCAATTGATAGGATGATGGGTTATAGACTAGTATCTAATCAG 1860

Qy   1861 gcaattgttagatgggtgttctctccagaaaatgttgatcagtttcatgtgtctgatcag 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GCAATTGTTAGATGGGTGTTCTCTCCAGAAAATGTTGATCAGTTTCATGTGTCTGATCAG 1920

Qy   1921 ccatgggagatacttggcaatgctcttaacaagacttataaccgtatctctgatttgagg 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CCATGGGAGATACTTGGCAATGCTCTTAACAAGACTTATAACCGTATCTCTGATTTGAGG 1980

Qy   1981 aaagatatatcaaacattacgaaaaatgttttggttgctgagaaagcttcagccaatgca 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AAAGATATATCAAACATTACGAAAAATGTTTTGGTTGCTGAGAAAGCTTCAGCCAATGCA 2040

Qy   2041 cgagtagagttggaggctgctgagagcaaactttccctagtggaaggtgaacccgttctt 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CGAGTAGAGTTGGAGGCTGCTGAGAGCAAACTTTCCCTAGTGGAAGGTGAACCCGTTCTT 2100

Qy   2101 ggtgagaatccagcgaagatgaagcgtttaaaatcaacagtggagaagacaggggaagcg 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GGTGAGAATCCAGCGAAGATGAAGCGTTTAAAATCAACAGTGGAGAAGACAGGGGAAGCG 2160

Qy   2161 gagttatctcttcgggagtccctagaggcaaaagaggctcttcttaacagagctctctct 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 GAGTTATCTCTTCGGGAGTCCCTAGAGGCAAAAGAGGCTCTTCTTAACAGAGCTCTCTCT 2220

Qy   2221 gagaccgaggttttactgctcttgctgttccaaagtttcttaggtgtactgaaggaacgg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GAGACCGAGGTTTTACTGCTCTTGCTGTTCCAAAGTTTCTTAGGTGTACTGAAGGAACGG 2280

Qy   2281 ctcccagatccaactaaagtgagatcagtgcaggatctaaaatctataggtgctgaagat 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CTCCCAGATCCAACTAAAGTGAGATCAGTGCAGGATCTAAAATCTATAGGTGCTGAAGAT 2340

Qy   2341 gacaagccatctgcgatggacgtggacagcgagaatggaaacccaaagaagagttgcgaa 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GACAAGCCATCTGCGATGGACGTGGACAGCGAGAATGGAAACCCAAAGAAGAGTTGCGAA 2400

Qy   2401 gtcggtgagagagaacagtggtgcttatcaacacttggctatctcacggcatttacaagg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GTCGGTGAGAGAGAACAGTGGTGCTTATCAACACTTGGCTATCTCACGGCATTTACAAGG 2460

Qy   2461 caatatgcgagcgagatatggcctcacatggagaagttggagtcagaagtgttctcgggt 2520
```

```
Db    2461 CAATATGCGAGCGAGATATGGCCTCACATGGAGAAGTTGGAGTCAGAAGTGTTCTCGGGT 2520

Qy    2521 gaagatgtgcatcctctctttctccaagccatatcttctgcacttcaattcccattacat 2580
Db    2521 GAAGATGTGCATCCTCTCTTTCTCCAAGCCATATCTTCTGCACTTCAATTCCCATTACAT 2580

Qy    2581 taatcttcctctttcaatctcaatcaaacctgtctcttttgtttttttgttatgagattct 2640
Db    2581 TAATCTTCCTCTTTCAATCTCAATCAAACCTGTCTCTTTTGTTTTTTGTTATGAGATTCT 2640

Qy    2641 gattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaagagtttaaacttt 2700
Db    2641 GATTCTGACATCAAGTTATTAGGAAATTGAAAAGAGTCAAAAAACAAGAGTTTAAACTTT 2700

Qy    2701 aaaaaaaaaaaaaaaa 2716
Db    2701 AAAAAAAAAAAAAAAA 2716


RESULT  3
AF268377
LOCUS       AF268377               2834 bp    mRNA    linear   PLN 14-JUN-2000
DEFINITION  Arabidopsis thaliana nuclear cap-binding protein CBP80 mRNA,
            complete cds.
ACCESSION   AF268377
VERSION     AF268377.1  GI:8515770
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2834)
  AUTHORS   Kmieciak,M. and Jarmolowski,A.
  TITLE     A nuclear cap-binding protein complex from Arabidopsis thaliana.
            Molecular characterization of Arabidopsis thaliana CBP80
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2834)
  AUTHORS   Kmieciak,M. and Jarmolowski,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Gene Expression, A. Mickiewicz University,
            Institute of Molecular Biology and Biotechnology, Miedzychodzka 5,
            Poznan 60-371, Poland
FEATURES             Location/Qualifiers
     source          1. .2834
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /dev_stage="ten day old"
                     /note="whole plant"
     CDS             170. .2716
                     /note="similar to Homo sapiens CBP80"
                     /codon_start=1
                     /product="nuclear cap-binding protein CBP80"
                     /protein_id="AAF76167.1"
                     /db_xref="GI:8515771"
                     /translation="MSNWKTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGD
                     QVLPFLLQCAEQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSI
                     RILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILS
                     SLPWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVE
                     DLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYS
                     GKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASY
                     MANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAV
                     RALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEIL
                     QREVRLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKV
                     KEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSK
                     LCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFH
                     VSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLV
                     EGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLFQS
                     FLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLS
                     TLGYLTAFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAISSALQFPLH"

BASE COUNT      823 a    553 c    650 g    808 t
ORIGIN


  Query Match             99.4%;  Score 2700;  DB 8;  Length 2834;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2700;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 aaagagacgaactgaagaaaaacctctcggaagaagatgagcaattggaaaactcttctc 60
Db     134 AAAGAGACGAACTGAAGAAAAACCTCTCGGAAGAAGATGAGCAATTGGAAAACTCTTCTC 193

Qy      61 cttcgcatcggcgaaaagggacctgagtacggcacttcctccgactacaaagaccacatc 120
Db     194 CTTCGCATCGGCGAAAAGGGACCTGAGTACGGCACTTCCTCCGACTACAAAGACCACATC 253

Qy     121 gagacttgtttcggtgtcattcgtagagaaatcgagcgttctggagatcaagttttgcct 180
Db     254 GAGACTTGTTTCGGTGTCATTCGTAGAGAAATCGAGCGTTCTGGAGATCAAGTTTTGCCT 313

Qy     181 tttctactacaatgtgctgaacaattgcctcataagattcctttgtatgggactttgatt 240
Db     314 TTTCTACTACAATGTGCTGAACAATTGCCTCATAAGATTCCTTTGTATGGGACTTTGATT 373

Qy     241 ggtttgttgaacttggagaatgaagattttgtccagaagctagtagaaagtgtccacgct 300
Db     374 GGTTTGTTGAACTTGGAGAATGAAGATTTTGTCCAGAAGCTAGTAGAAAGTGTCCACGCT 433

Qy     301 aatttccaggtcgctttagattctggcaactgcaacagtatccgtatattgcttcgcttt 360
Db     434 AATTTCCAGGTCGCTTTAGATTCTGGCAACTGCAACAGTATCCGTATATTGCTTCGCTTT 493

Qy     361 atgacttccctgttgtgcagtaaggttattcaacctgcttctttgattgtcgtcttcgaa 420
Db     494 ATGACTTCCCTGTTGTGCAGTAAGGTTATTCAACCTGCTTCTTTGATTGTCGTCTTCGAA 553

Qy     421 acattgctatcatctgctgccactactgtggatgaagagaaaggaaatccatcatggcag 480
Db     554 ACATTGCTATCATCTGCTGCCACTACTGTGGATGAAGAGAAAGGAAATCCATCATGGCAG 613

Qy     481 ccacaagctgacttttacgttatatgcatcttgtccagcctcccgtggggaggatcagaa 540
Db     614 CCACAAGCTGACTTTTACGTTATATGCATCTTGTCCAGCCTCCCGTGGGGAGGATCAGAA 673

Qy     541 ctcgctgagcaagttcctgatgagattgaaagagtgttagttgggatacaagcttatttg 600
Db     674 CTCGCTGAGCAAGTTCCTGATGAGATTGAAAGAGTGTTAGTTGGGATACAAGCTTATTTG 733

Qy     601 agcatccgaaagaattcttccacctctgggttaaacttttttcacaacggagaatttgaa 660
Db     734 AGCATCCGAAAGAATTCTTCCACCTCTGGGTTAAACTTTTTTCACAACGGAGAATTTGAA 793

Qy     661 agcagccttgcagagaaggatttcgtggaggatctattggatcgaattcagtctctggct 720
Db     794 AGCAGCCTTGCAGAGAAGGATTTCGTGGAGGATCTATTGGATCGAATTCAGTCTCTGGCT 853

Qy     721 tccaatggatggaaacttgaaagcgtacctaggcctcatctctcgtttgaagctcaactc 780
Db     854 TCCAATGGATGGAAACTTGAAAGCGTACCTAGGCCTCATCTCTCGTTTGAAGCTCAACTC 913

Qy     781 gttgctggaaagtttcatgagctacgtcccattaaatgtatggaacaaccgagtccacct 840
Db     914 GTTGCTGGAAAGTTTCATGAGCTACGTCCCATTAAATGTATGGAACAACCGAGTCCACCT 973

Qy     841 tctgatcattcgagggcatacagtggcaagcaaaagcatgatgcattgacgagatatccc 900
Db     974 TCTGATCATTCGAGGGCATACAGTGGCAAGCAAAAGCATGATGCATTGACGAGATATCCC 1033

Qy     901 cagagaattcgtaggttgaatatatttccagctaataaaatggaggatgtacaaccaatt 960
Db    1034 CAGAGAATTCGTAGGTTGAATATATTTCCAGCTAATAAAATGGAGGATGTACAACCAATT 1093

Qy     961 gatcgttttgtcgtggaggagtatttgctggatgtgctcttctatttgaatggatgtcgg 1020
```

```
Db    1094 GATCGTTTTGTCGTGGAGGAGTATTTGCTGGATGTGCTCTTCTATTTGAATGGATGTCGG 1153

Qy    1021 aaggagtgtgcatcctacatggctaatcttcctgttacatttcggtacgagtatcttatg 1080
Db    1154 AAGGAGTGTGCATCCTACATGGCTAATCTTCCTGTTACATTTCGGTACGAGTATCTTATG 1213

Qy    1081 gcagagacactatttttctcagatactgctgctaccccagccaccattcaagactctttat 1140
Db    1214 GCAGAGACACTATTTTCTCAGATACTGCTGCTACCCCAGCCACCATTCAAGACTCTTTAT 1273

Qy    1141 tatacactcgtgattatggatctttgtaaggctcttccgggtgcctttcctgctgttgtt 1200
Db    1274 TATACACTCGTGATTATGGATCTTTGTAAGGCTCTTCCGGGTGCCTTTCCTGCTGTTGTT 1333

Qy    1201 gctggcgctgttcgtgcactatttgagaaaatatccgacttagacatggaatccaggacg 1260
Db    1334 GCTGGCGCTGTTCGTGCACTATTTGAGAAAATATCCGACTTAGACATGGAATCCAGGACG 1393

Qy    1261 cgtcttatcctctggttttctcaccacttatccaacttccaattcatctggccgtgggaa 1320
Db    1394 CGTCTTATCCTCTGGTTTTCTCACCACTTATCCAACTTCCAATTCATCTGGCCGTGGGAA 1453

Qy    1321 gagtgggcttttgtgttggatcttcccaagtgggcccctaagcgtgtatttgttcaggag 1380
Db    1454 GAGTGGGCTTTTGTGTTGGATCTTCCCAAGTGGGCCCCTAAGCGTGTATTTGTTCAGGAG 1513

Qy    1381 attttgcaaagagaagtacgcttgtcttactgggataaaattaagcagagcattgagaat 1440
Db    1514 ATTTTGCAAAGAGAAGTACGCTTGTCTTACTGGGATAAAATTAAGCAGAGCATTGAGAAT 1573

Qy    1441 gcgactgccctagaagaattacttcctccaaaaagctggtccgaattttatgtattccttg 1500
Db    1574 GCGACTGCCCTAGAAGAATTACTTCCTCCAAAAGCTGGTCCGAATTTTATGTATTCCTTG 1633

Qy    1501 gaagaaggtaaagagaaaacagaagaacagcaattgtcagccgaattgagcaggaaggtc 1560
Db    1634 GAAGAAGGTAAAGAGAAAACAGAAGAACAGCAATTGTCAGCCGAATTGAGCAGGAAGGTC 1693

Qy    1561 aaggaaaaacaaaccgcacgtgacatgatagtgtggattgaagaaacgatatatccagtt 1620
Db    1694 AAGGAAAAACAAACCGCACGTGACATGATAGTGTGGATTGAAGAAACGATATATCCAGTT 1753

Qy    1621 catggttttgaagttactcttacaatagttgtacagaccttacttgacatcggatcaaaa 1680
Db    1754 CATGGTTTTGAAGTTACTCTTACAATAGTTGTACAGACCTTACTTGACATCGGATCAAAA 1813

Qy    1681 agtttcactcatttggtcactgtcctggagcgatatggccaagtattttcaaagctttgt 1740
Db    1814 AGTTTCACTCATTTGGTCACTGTCCTGGAGCGATATGGCCAAGTATTTTCAAAGCTTTGT 1873

Qy    1741 cctgataacgataagcaggtgatgctattatctcaagtgagtacatactggaaaaacaat 1800
Db    1874 CCTGATAACGATAAGCAGGTGATGCTATTATCTCAAGTGAGTACATACTGGAAAAACAAT 1933

Qy    1801 gtacaaatgacggcggtggcaattgataggatgatgggttatagactagtatctaatcag 1860
Db    1934 GTACAAATGACGGCGGTGGCAATTGATAGGATGATGGGTTATAGACTAGTATCTAATCAG 1993

Qy    1861 gcaattgttagatgggtgttctctccagaaaatgttgatcagtttcatgtgtctgatcag 1920
Db    1994 GCAATTGTTAGATGGGTGTTCTCTCCAGAAAATGTTGATCAGTTTCATGTGTCTGATCAG 2053

Qy    1921 ccatgggagatacttggcaatgctcttaacaagacttataaccgtatctctgatttgagg 1980
Db    2054 CCATGGGAGATACTTGGCAATGCTCTTAACAAGACTTATAACCGTATCTCTGATTTGAGG 2113

Qy    1981 aaagatatatcaaacattacgaaaaatgttttggttgctgagaaagcttcagccaatgca 2040
Db    2114 AAAGATATATCAAACATTACGAAAAATGTTTTGGTTGCTGAGAAAGCTTCAGCCAATGCA 2173

Qy    2041 cgagtagagttggaggctgctgagagcaaactttccctagtggaaggtgaacccgttctt 2100
Db    2174 CGAGTAGAGTTGGAGGCTGCTGAGAGCAAACTTTCCCTAGTGGAAGGTGAACCCGTTCTT 2233

Qy    2101 ggtgagaatccagcgaagatgaagcgtttaaaatcaacagtggagaagacaggggaagcg 2160
Db    2234 GGTGAGAATCCAGCGAAGATGAAGCGTTTAAAATCAACAGTGGAGAAGACAGGGGAAGCG 2293

Qy    2161 gagttatctcttcgggagtccctagaggcaaaagaggctcttcttaacagagctctctct 2220
Db    2294 GAGTTATCTCTTCGGGAGTCCCTAGAGGCAAAAGAGGCTCTTCTTAACAGAGCTCTCTCT 2353

Qy    2221 gagaccgaggttttactgctcttgctgttccaaagtttcttaggtgtactgaaggaacgg 2280
Db    2354 GAGACCGAGGTTTTACTGCTCTTGCTGTTCCAAAGTTTCTTAGGTGTACTGAAGGAACGG 2413

Qy    2281 ctcccagatccaactaaagtgagatcagtgcaggatctaaaatctataggtgctgaagat 2340
Db    2414 CTCCCAGATCCAACTAAAGTGAGATCAGTGCAGGATCTAAAATCTATAGGTGCTGAAGAT 2473

Qy    2341 gacaagccatctgcgatggacgtggacagcgagaatggaaacccaaagaagagttgcgaa 2400
Db    2474 GACAAGCCATCTGCGATGGACGTGGACAGCGAGAATGGAAACCCAAAGAAGAGTTGCGAA 2533

Qy    2401 gtcggtgagagagaacagtggtgcttatcaacacttggctatctcacggcatttacaagg 2460
Db    2534 GTCGGTGAGAGAGAACAGTGGTGCTTATCAACACTTGGCTATCTCACGGCATTTACAAGG 2593

Qy    2461 caatatgcgagcgagatatggcctcacatggagaagttggagtcagaagtgttctcgggt 2520
Db    2594 CAATATGCGAGCGAGATATGGCCTCACATGGAGAAGTTGGAGTCAGAAGTGTTCTCGGGT 2653

Qy    2521 gaagatgtgcatcctctctttctccaagccatatcttctgcacttcaattcccattacat 2580
Db    2654 GAAGATGTGCATCCTCTCTTTCTCCAAGCCATATCTTCTGCACTTCAATTCCCATTACAT 2713

Qy    2581 taatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgttatgagattct 2640
Db    2714 TAATCTTCCTCTTTCAATCTCAATCAAACCTGTCTCTTTTGTTTTTTGTTATGAGATTCT 2773

Qy    2641 gattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaagagtttaaacttt 2700
Db    2774 GATTCTGACATCAAGTTATTAGGAAATTGAAAAGAGTCAAAAAACAAGAGTTTAAACTTT 2833
```

```
RESULT   4
AY064960
LOCUS       AY064960               2735 bp    mRNA    linear   PLN 10-DEC-2001
DEFINITION  Arabidopsis thaliana At2g13540/T10F5.8 mRNA, complete cds.
ACCESSION   AY064960
VERSION     AY064960.1  GI:17473508
KEYWORDS    FLI_CDNA.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2735)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L.,
            Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K.,
            Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A.
            and Ecker,J.R.
  TITLE     Arabidopsis cDNA clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2735)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L.,
            Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K.,
```

```
            Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A.
            and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-2001) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H.,
            Kim,C.J., Meyers,M.C., Shinn,P., Banh,J. Bowser,L., Dale,J.M.,
            Goldsmith,A.D., Jiang,P.X., Jones,T., Karlin-Neumann,G., Lam,B.,
            Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M., Onodera,C.S.,
            Palm,C.J., Pham,P.K., Quach,H.L., Southwick,A., Tang,C.C.,
            Toriumi,M., Yamada,K., Yamamura, Y., Yu,G., Yu,S., Davis,R.W.,
            Theologis,A., and Ecker,J.R.

            Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .2735
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /clone="RAFL09-32-L06(R21227)"
                     /note="ecotype: Columbia"
     5'UTR           1. .70
     CDS             71. .2617
                     /note="putative cap-binding protein"
                     /codon_start=1
                     /product="At2g13540/T10F5.8"
                     /protein_id="AAL38378.1"
                     /db_xref="GI:17473509"
                     /translation="MSNWKTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGD
                     QVLPFLLQCAEQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSI
                     RILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILS
                     SLPWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVE
                     DLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYS
                     GKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASY
                     MANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAV
                     RALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEIL
                     QREVRLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKV
                     KEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSK
                     LCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFH
                     MSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLV
                     EGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLFQS
                     FLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLS
                     TLGYLTAFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAISSALQFPLH"
     misc_difference 1943
                     /note="compared to genome"
                     /replace="g"
     3'UTR           2618. .2735
BASE COUNT      778 a    539 c    637 g    781 t
ORIGIN
```

Query Match 99.3%; Score 2697.4; DB 8; Length 2735;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2698; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 aaagagacgaactgaagaaaaacctctcggaagaagatgagcaattggaaaactcttctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35 AAAGAGACGAACTGAAGAAAAACCTCTCGGAAGAAGATGAGCAATTGGAAAACTCTTCTC 94

Qy      61 cttcgcatcggcgaaaagggacctgagtacggcacttcctccgactacaaagaccacatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      95 CTTCGCATCGGCGAAAAGGGACCTGAGTACGGCACTTCCTCCGACTACAAAGACCACATC 154

Qy     121 gagacttgtttcggtgtcattcgtagagaaatcgagcgttctggagatcaagttttgcct 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155 GAGACTTGTTTCGGTGTCATTCGTAGAGAAATCGAGCGTTCTGGAGATCAAGTTTTGCCT 214

Qy     181 tttctactacaatgtgctgaacaattgcctcataagattcctttgtatgggactttgatt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     215 TTTCTACTACAATGTGCTGAACAATTGCCTCATAAGATTCCTTTGTATGGGACTTTGATT 274

Qy     241 ggtttgttgaacttggagaatgaagattttgtccagaagctagtagaaagtgtccacgct 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     275 GGTTTGTTGAACTTGGAGAATGAAGATTTTGTCCAGAAGCTAGTAGAAAGTGTCCACGCT 334

Qy     301 aatttccaggtcgctttagattctggcaactgcaacagtatccgtatattgcttcgcttt 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     335 AATTTCCAGGTCGCTTTAGATTCTGGCAACTGCAACAGTATCCGTATATTGCTTCGCTTT 394

Qy     361 atgacttccctgttgtgcagtaaggttattcaacctgcttctttgattgtcgtcttcgaa 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     395 ATGACTTCCCTGTTGTGCAGTAAGGTTATTCAACCTGCTTCTTTGATTGTCGTCTTCGAA 454

Qy     421 acattgctatcatctgctgccactactgtggatgaagagaaaggaaatccatcatggcag 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     455 ACATTGCTATCATCTGCTGCCACTACTGTGGATGAAGAGAAAGGAAATCCATCATGGCAG 514

Qy     481 ccacaagctgacttttacgttatatgcatcttgtccagcctcccgtggggaggatcagaa 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     515 CCACAAGCTGACTTTTACGTTATATGCATCTTGTCCAGCCTCCCGTGGGGAGGATCAGAA 574

Qy     541 ctcgctgagcaagttcctgatgagattgaaagagtgttagttgggatacaagcttatttg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     575 CTCGCTGAGCAAGTTCCTGATGAGATTGAAAGAGTGTTAGTTGGGATACAAGCTTATTTG 634

Qy     601 agcatccgaaagaattcttccacctctgggttaaacttttttcacaacggagaatttgaa 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     635 AGCATCCGAAAGAATTCTTCCACCTCTGGGTTAAACTTTTTTCACAACGGAGAATTTGAA 694

Qy     661 agcagccttgcagagaaggatttcgtggaggatctattggatcgaattcagtctctggct 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     695 AGCAGCCTTGCAGAGAAGGATTTCGTGGAGGATCTATTGGATCGAATTCAGTCTCTGGCT 754

Qy     721 tccaatggatggaaacttgaaagcgtacctaggcctcatctctcgtttgaagctcaactc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     755 TCCAATGGATGGAAACTTGAAAGCGTACCTAGGCCTCATCTCTCGTTTGAAGCTCAACTC 814

Qy     781 gttgctggaaagtttcatgagctacgtcccattaaatgtatggaacaaccgagtccacct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     815 GTTGCTGGAAAGTTTCATGAGCTACGTCCCATTAAATGTATGGAACAACCGAGTCCACCT 874

Qy     841 tctgatcattcgagggcatacagtggcaagcaaaagcatgatgcattgacgagatatccc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     875 TCTGATCATTCGAGGGCATACAGTGGCAAGCAAAAGCATGATGCATTGACGAGATATCCC 934

Qy     901 cagagaattcgtaggttgaatatatttccagctaataaaatggaggatgtacaaccaatt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     935 CAGAGAATTCGTAGGTTGAATATATTTCCAGCTAATAAAATGGAGGATGTACAACCAATT 994

Qy     961 gatcgttttgtcgtggaggagtatttgctggatgtgctcttctatttgaatggatgtcgg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     995 GATCGTTTTGTCGTGGAGGAGTATTTGCTGGATGTGCTCTTCTATTTGAATGGATGTCGG 1054

Qy    1021 aaggagtgtgcatcctacatggctaatcttcctgttacatttcggtacgagtatcttatg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1055 AAGGAGTGTGCATCCTACATGGCTAATCTTCCTGTTACATTTCGGTACGAGTATCTTATG 1114

Qy    1081 gcagagacactatttttctcagatactgctgctaccccagccaccattcaagactctttat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1115 GCAGAGACACTATTTTCTCAGATACTGCTGCTACCCCAGCCACCATTCAAGACTCTTTAT 1174

Qy    1141 tatacactcgtgattatggatctttgtaaggctcttccgggtgcctttcctgctgttgtt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1175 TATACACTCGTGATTATGGATCTTTGTAAGGCTCTTCCGGGTGCCTTTCCTGCTGTTGTT 1234
```

```
Qy   1201 gctggcgctgttcgtgcactatttgagaaaatatccgacttagacatggaatccaggacg 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1235 GCTGGCGCTGTTCGTGCACTATTTGAGAAAATATCCGACTTAGACATGGAATCCAGGACG 1294

Qy   1261 cgtcttatcctctggttttctcaccacttatccaacttccaattcatctggccgtgggaa 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1295 CGTCTTATCCTCTGGTTTTCTCACCACTTATCCAACTTCCAATTCATCTGGCCGTGGGAA 1354

Qy   1321 gagtgggcttttgtgttggatcttcccaagtgggcccctaagcgtgtatttgttcaggag 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1355 GAGTGGGCTTTTGTGTTGGATCTTCCCAAGTGGGCCCCTAAGCGTGTATTTGTTCAGGAG 1414

Qy   1381 attttgcaaagagaagtacgcttgtcttactgggataaaattaagcagagcattgagaat 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1415 ATTTTGCAAAGAGAAGTACGCTTGTCTTACTGGGATAAAATTAAGCAGAGCATTGAGAAT 1474

Qy   1441 gcgactgccctagaagaattacttcctccaaaagctggtccgaattttatgtattccttg 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1475 GCGACTGCCCTAGAAGAATTACTTCCTCCAAAAGCTGGTCCGAATTTTATGTATTCCTTG 1534

Qy   1501 gaagaaggtaaagagaaaacagaagaacagcaattgtcagccgaattgagcaggaaggtc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1535 GAAGAAGGTAAAGAGAAAACAGAAGAACAGCAATTGTCAGCCGAATTGAGCAGGAAGGTC 1594

Qy   1561 aaggaaaaacaaaccgcacgtgacatgatagtgtggattgaagaaacgatatatccagtt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1595 AAGGAAAAACAAACCGCACGTGACATGATAGTGTGGATTGAAGAAACGATATATCCAGTT 1654

Qy   1621 catggttttgaagttactcttacaatagttgtacagaccttacttgacatcggatcaaaa 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1655 CATGGTTTTGAAGTTACTCTTACAATAGTTGTACAGACCTTACTTGACATCGGATCAAAA 1714

Qy   1681 agtttcactcatttggtcactgtcctggagcgatatggccaagtattttcaaagctttgt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1715 AGTTTCACTCATTTGGTCACTGTCCTGGAGCGATATGGCCAAGTATTTTCAAAGCTTTGT 1774

Qy   1741 cctgataacgataagcaggtgatgctattatctcaagtgagtacatactggaaaaacaat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1775 CCTGATAACGATAAGCAGGTGATGCTATTATCTCAAGTGAGTACATACTGGAAAAACAAT 1834

Qy   1801 gtacaaatgacggcggtggcaattgataggatgatgggttatagactagtatctaatcag 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1835 GTACAAATGACGGCGGTGGCAATTGATAGGATGATGGGTTATAGACTAGTATCTAATCAG 1894

Qy   1861 gcaattgttagatgggtgttctctccagaaaatgttgatcagtttcatgtgtctgatcag 1920
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   1895 GCAATTGTTAGATGGGTGTTCTCTCCAGAAAATGTTGATCAGTTTCATATGTCTGATCAG 1954

Qy   1921 ccatgggagatacttggcaatgctcttaacaagacttataaccgtatctctgatttgagg 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1955 CCATGGGAGATACTTGGCAATGCTCTTAACAAGACTTATAACCGTATCTCTGATTTGAGG 2014

Qy   1981 aaagatatatcaaacattacgaaaaatgttttggttgctgagaaagcttcagccaatgca 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2015 AAAGATATATCAAACATTACGAAAAATGTTTTGGTTGCTGAGAAAGCTTCAGCCAATGCA 2074

Qy   2041 cgagtagagttggaggctgctgagagcaaactttccctagtggaaggtgaacccgttctt 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2075 CGAGTAGAGTTGGAGGCTGCTGAGAGCAAACTTTCCCTAGTGGAAGGTGAACCCGTTCTT 2134

Qy   2101 ggtgagaatccagcgaagatgaagcgtttaaaatcaacagtggagaagacaggggaagcg 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2135 GGTGAGAATCCAGCGAAGATGAAGCGTTTAAAATCAACAGTGGAGAAGACAGGGGAAGCG 2194

Qy   2161 gagttatctcttcgggagtccctagaggcaaaagaggctcttcttaacagagctctctct 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2195 GAGTTATCTCTTCGGGAGTCCCTAGAGGCAAAAGAGGCTCTTCTTAACAGAGCTCTCTCT 2254

Qy   2221 gagaccgaggttttactgctcttgctgttccaaagtttcttaggtgtactgaaggaacgg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2255 GAGACCGAGGTTTTACTGCTCTTGCTGTTCCAAAGTTTCTTAGGTGTACTGAAGGAACGG 2314

Qy   2281 ctcccagatccaactaaagtgagatcagtgcaggatctaaaatctataggtgctgaagat 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2315 CTCCCAGATCCAACTAAAGTGAGATCAGTGCAGGATCTAAAATCTATAGGTGCTGAAGAT 2374

Qy   2341 gacaagccatctgcgatggacgtggacagcgagaatggaaacccaaagaagagttgcgaa 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2375 GACAAGCCATCTGCGATGGACGTGGACAGCGAGAATGGAAACCCAAAGAAGAGTTGCGAA 2434

Qy   2401 gtcggtgagagagaacagtggtgcttatcaacacttggctatctcacggcatttacaagg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2435 GTCGGTGAGAGAGAACAGTGGTGCTTATCAACACTTGGCTATCTCACGGCATTTACAAGG 2494

Qy   2461 caatatgcgagcgagatatggcctcacatggagaagttggagtcagaagtgttctcgggt 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2495 CAATATGCGAGCGAGATATGGCCTCACATGGAGAAGTTGGAGTCAGAAGTGTTCTCGGGT 2554

Qy   2521 gaagatgtgcatcctctctttctccaagccatatcttctgcacttcaattcccattacat 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2555 GAAGATGTGCATCCTCTCTTTCTCCAAGCCATATCTTCTGCACTTCAATTCCCATTACAT 2614

Qy   2581 taatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgttatgagattct 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2615 TAATCTTCCTCTTTCAATCTCAATCAAACCTGTCTCTTTTGTTTTTTGTTATGAGATTCT 2674

Qy   2641 gattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaagagtttaaacttt 2699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2675 GATTCTGACATCAAGTTATTAGGAAATTGAAAAGAGTCAAAAAACAAGAGTTTAAACTT 2733
```

RESULT 5
AC007063
LOCUS AC007063 82403 bp DNA linear PLN 05-APR-2000
DEFINITION Arabidopsis thaliana chromosome II section 76 of 255 of the complete sequence. Sequence from clones T10F5, F13J11.
ACCESSION AC007063 AE002093
VERSION AC007063.5 GI:6598657
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 82403)
AUTHORS Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D., Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E., Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M., Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L., Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H., Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D., Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and Venter,J.C.
TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20083487
PUBMED 10617197
REFERENCE 2 (bases 1 to 82403)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Dec 17, 1999 this sequence version replaced gi:4558656.
The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), searches of the complete sequence against a peptide database and plant EST

(026) 4/2/99

databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES Location/Qualifiers

source 1. .82403
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"

misc_feature 1. .82202
/note="Sequence from clone T10F5"

repeat_region 1197. .1306
/rpt_family="(TAAAA)n"

repeat_region complement(3226. .3308)
/rpt_family="(TAAAA)n"

repeat_region 7744. .7865
/rpt_family="(TAAAA)n"

gene complement(7998. .11670)
/gene="At2g13470"
/note="T10F5.1"
/pseudo

gene complement(12098. .12827)
/gene="At2g13480"
/note="T10F5.2; similar to GB:AAA75253"
/pseudo

repeat_region complement(13418. .13471)
/rpt_family="(CAAAA)n"

gene 14215. .18083
/gene="At2g13490"
/note="T10F5.3"
/pseudo

repeat_region complement(18097. .18209)
/rpt_family="(CAAAA)n"

mRNA join(<18363. .19150,19298. .>19526)
/gene="At2g13500"

gene <18363. .>19526
/gene="At2g13500"
/note="T10F5.4; similar to GB:AAA75253; some members of this protein family have a weak CCHC zinc fingers that is mostly from retroviral gag proteins (nucleocapsid)"

CDS join(18363. .19150,19298. .19526)
/gene="At2g13500"
/codon_start=1
/product="putative Tal1-like non-LTR retroelement protein"
/protein_id="AAD22675.1"
/db_xref="GI:4558657"
/translation="MVDVLWDEIQNLELGQDYPTLFIPNEAYVMVEDRNRLSLIARPL
LPRVQNFNAVINALPRAWGLTAHVHGRILDATYVQLLFQTEIYLLSVQRREVWLFNNW
FVASQRWEAAPSLNFVTTIDLWVQMRGIPILYVFEKTELEIAQELGEIISLDFHDATS
TQIAYIRVRVRFGITDRLRFFQRITFESGETALIRFQYERLRRIYSCFRFTHNLNYYP
YRQRFQSIDRERAVFRDSVLRSSFNSQSQMTDSSLQVPLHPPPRVPSSGDAITSGIMR
VFEVGESSRGVQDGESHNNMNLGESSKRKNLVILQRDKESRQHKQQETEQRMGGILKP
PKKR"

mRNA complement(join(<20428. .20746,20999. .>21534))
/gene="At2g13510"

gene complement(<20428. .>21534)
/gene="At2g13510"
/note="T10F5.5; similar to GB:AAA75253; some members of this protein family have a weak CCHC zinc fingers that is mostly from retroviral gag proteins (nucleocapsid)"

CDS complement(join(20428. .20746,20999. .21534))
/gene="At2g13510"
/codon_start=1
/product="putative Tal1-like non-LTR retroelement protein"
/protein_id="AAD22676.1"
/db_xref="GI:4558658"
/translation="MADKLWDEIQNLELGPEDPTLFIPHEAYVMVAETNRLSIIARPL
NPRVQNLYSVIAALPRSWGLTTHVHGRVLDATYVQFLFHTEMDLVSVQRREPWLFNNR
FVASQRWEPAPALNFVTTIDLLVQMRGIPLPYVYEETALEIAQEIGEIISLDFHDATS
TQIAYIRVRVQIGITERLRVPAPPLNPEELAGASPYFPSSRNAAFQPYMVPSTQGGVG
RHHVFPDSNINPSSGAAITSGVRRVYEVGESFRPCPDGESRNNRNIGESSKRKNLVIL
QSDNEKAN"

repeat_region 21554. .21667
/rpt_family="(TAAAA)n"

gene complement(21749. .25813)
/gene="At2g13520"
/note="T10F5.6"
/pseudo

gene complement(25902. .26957)
/gene="At2g13530"
/note="T10F5.7"
/pseudo

mRNA join(<27480. .27563,27648. .27707,27807. .27863,28349. .28420,
28529. .28591,28765. .28941,29126. .29254,29478. .29544,
29724. .29923,30139. .30209,30672. .30756,30903. .30971,
31063. .31181,31489. .31627,31829. .32329,32420. .32778,
33154. .33372)
/gene="At2g13540"

gene <27480. .33372
/gene="At2g13540"
/note="T10F5.8"

CDS join(27480. .27563,27648. .27707,27807. .27863,28349. .28420,
28529. .28591,28765. .28941,29126. .29254,29478. .29544,
29724. .29923,30139. .30209,30672. .30756,30903. .30971,
31063. .31181,31489. .31627,31829. .32329,32420. .32776)
/gene="At2g13540"
/codon_start=1
/product="putative cap-binding protein"
/protein_id="AAD22677.1"
/db_xref="GI:4558659"
/translation="MSNWKTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGD
QVLPFLLQCAEQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSI
RILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILS
SLPWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVE
DLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYS
GKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASY
MANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAV
RALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEIL
QREVRLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKV
KEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSK
LCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFH
VSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLV
EGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLNRALSETEVLYISLMQH
MILLFPFIY"

repeat_region complement(34637. .34727)
/rpt_family="(TA)n"

mRNA complement(38932. .40224)
/gene="At2g13550"

gene complement(38932. .40224)
/gene="At2g13550"
/note="T10F5.9"

CDS complement(39231. .39776)

```
                     /gene="At2g13550"
                     /note="unknown protein"
                     /codon_start=1
                     /protein_id="AAD22678.1"
                     /db_xref="GI:4558660"
                     /translation="MTSQIRQTPLARMHLPTQFQPNTRTGRQPKSPPNSHHPDESSPS
                     PQQERHITRSEPITRGIHQLKTTSSSTANTQNTTGEDKRKQFHIVTNTDIASTDKTRS
                     RSGNQSKPQPHIQPKLKQSLKTGAANNIETNLRNRYRREIGPRTPQHHQPYNHITIKT
                     RRKKSTDEIKQRRTRPESVLR"
     mRNA            join(<40516. .40686,40766. .40862,40943. .41061,41171. .41338,
                     41499. .41627,41718. .41778,42210. .42313,42390. .42473,

  Query Match             36.5%;  Score 990.2;  DB 8;  Length 82403;
  Best Local Similarity   83.1%;  Pred. No. 3.4e-232;
  Matches 1279;  Conservative    0;  Mismatches    3;  Indels  257;  Gaps    3;

Qy    1424 agcagagcattgagaatgcgactgccctagaagaattacttcctccaaaagctggtccga 1483
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31824 AACAGAGCATTGAGAATGCGACTGCCCTAGAAGAATTACTTCCTCCAAAAGCTGGTCCGA 31883

Qy    1484 attttatgtattccttggaagaaggtaaagagaaaacagaagaacagcaattgtcagccg 1543
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31884 ATTTTATGTATTCCTTGGAAGAAGGTAAAGAGAAAACAGAAGAACAGCAATTGTCAGCCG 31943

Qy    1544 aattgagcaggaaggtcaaggaaaaacaaaccgcacgtgacatgatagtgtggattgaag 1603
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31944 AATTGAGCAGGAAGGTCAAGGAAAAACAAACCGCACGTGACATGATAGTGTGGATTGAAG 32003

Qy    1604 aaacgatatatccagttcatggttttgaagttactcttacaatagttgtacagaccttac 1663
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32004 AAACGATATATCCAGTTCATGGTTTTGAAGTTACTCTTACAATAGTTGTACAGACCTTAC 32063

Qy    1664 ttgacatcggatcaaaaagtttcactcatttggtcactgtcctggagcgatatggccaag 1723
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32064 TTGACATCGGATCAAAAAGTTTCACTCATTTGGTCACTGTCCTGGAGCGATATGGCCAAG 32123

Qy    1724 tattttcaaagctttgtcctgataacgataagcaggtgatgctattatctcaagtgagta 1783
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32124 TATTTTCAAAGCTTTGTCCTGATAACGATAAGCAGGTGATGCTATTATCTCAAGTGAGTA 32183

Qy    1784 catactggaaaaacaatgtacaaatgacggcggtggcaattgataggatgatgggttata 1843
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32184 CATACTGGAAAAACAATGTACAAATGACGGCGGTGGCAATTGATAGGATGATGGGTTATA 32243

Qy    1844 gactagtatctaatcaggcaattgttagatgggtgttctctccagaaaatgttgatcagt 1903
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32244 GACTAGTATCTAATCAGGCAATTGTTAGATGGGTGTTCTCTCCAGAAAATGTTGATCAGT 32303

Qy    1904 ttcatgtgtctgatcagccatggg---------------------------------- 1927
           |||||||||||||||||||||||
Db   32304 TTCATGTGTCTGATCAGCCATGGGAGGTCAGTTGCGTTTTAACAGAAGTATATTTATTTT 32363

Qy    1928 ----------------------------------------------------agatac 1933
                                                               ||||||
Db   32364 CTACACTCTGCAGAGCAAGTTTGTTTATCTTCGTAAATCTGTTTCTGTCTCTGCAGATAC 32423

Qy    1934 ttggcaatgctcttaacaagacttataaccgtatctctgatttgaggaaagatatatcaa 1993
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32424 TTGGCAATGCTCTTAACAAGACTTATAACCGTATCTCTGATTTGAGGAAAGATATATCAA 32483

Qy    1994 acattacgaaaaatgttttggttgctgagaaagcttcagccaatgcacgagtagagttgg 2053
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32484 ACATTACGAAAAATGTTTTGGTTGCTGAGAAAGCTTCAGCCAATGCACGAGTAGAGTTGG 32543

Qy    2054 aggctgctgagagcaaactttccctagtggaaggtgaacccgttcttggtgagaatccag 2113
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32544 AGGCTGCTGAGAGCAAACTTTCCCTAGTGGAAGGTGAACCCGTTCTTGGTGAGAATCCAG 32603

Qy    2114 cgaagatgaagcgtttaaaatcaacagtggagaagacaggggaagcggagttatctcttc 2173
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32604 CGAAGATGAAGCGTTTAAAATCAACAGTGGAGAAGACAGGGGAAGCGGAGTTATCTCTTC 32663

Qy    2174 gggagtccctagaggcaaaagaggctcttcttaacagagctctctctgagaccg------ 2227
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32664 GGGAGTCCCTAGAGGCAAAAGAGGCTCTTCTTAACAGAGCTCTCTCTGAGACCGAGGTAT 32723

Qy    2228 ---------------------------------------------------------- 2227

Db   32724 TATATATTAGTTTAATGCAGCACATGATCTTATTATTTCCTTTTATCTACTAAAAGCTAA 32783

Qy    2228 ------------------aggttttactgctcttgctgttccaaagtttcttaggtgtac 2269
                             ||||||||||||||||||||||||||||||||||||||||||||
Db   32784 TTGAGTTGACTTGTGCATAGGTTTTACTGCTCTTGCTGTTCCAAAGTTTCTTAGGTGTAC 32843

Qy    2270 tgaaggaacggctcccagatccaactaaagtgagatcagtgcaggatctaaaatctatag 2329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32844 TGAAGGAACGGCTCCCAGATCCAACTAAAGTGAGATCAGTGCAGGATCTAAAATCTATAG 32903

Qy    2330 gtgctgaagatgacaagccatctgcgatggacgtggacagcgagaatggaaacccaaaga 2389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32904 GTGCTGAAGATGACAAGCCATCTGCGATGGACGTGGACAGCGAGAATGGAAACCCAAAGA 32963

Qy    2390 agagttgcgaagtcggtgagagagaacagtggtgcttatcaacacttggctatctcacgg 2449
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32964 AGAGTTGCGAAGTCGGTGAGAGAGAACAGTGGTGCTTATCAACACTTGGCTATCTCACGG 33023

Qy    2450 catttacaaggcaatatgcgagcg---------------------------------- 2473
           |||||||||||||||||||||||
Db   33024 CATTTACAAGGCAATATGCGAGCGAGGTAAAAAATAACTTAATAGTTGGGATGTTGTTGT 33083

Qy    2474 ----------------------------------------------agatatggcctca 2486
                                                           |||||||||||||
Db   33084 AAGAGCTTTAGTAGATTGAATATTGACTTGTTTCACCATGTTCTTGTAGATATGGCCTCA 33143

Qy    2487 catggagaagttggagtcagaagtgttctcgggtgaagatgtgcatcctctctttctcca 2546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33144 CATGGAGAAGTTGGAGTCAGAAGTGTTCTCGGGTGAAGATGTGCATCCTCTCTTTCTCCA 33203

Qy    2547 agccatatcttctgcacttcaattcccattacattaatcttcctctttcaatctcaatca 2606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33204 AGCCATATCTTCTGCACTTCAATTCCCATTACATTAATCTTCCTCTTTCAATCTCAATCA 33263

Qy    2607 aacctgtctcttttgtttttttgttatgagattctgattctgacatcaagttattaggaaa 2666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33264 AACCTGTCTCTTTTGTTTTTTGTTATGAGATTCTGATTCTGACATCAAGTTATTAGGAAA 33323

Qy    2667 ttgaaaagagtcaaaaaacaagagtttaaacttttaaaaa 2705
           |||||||||||||||||||||||||||||||| || |
Db   33324 TTGAAAAGAGTCAAAAAACAAGAGTTTAAACTTTTAACA 33362


RESULT   6
AY017415
LOCUS       AY017415               3009 bp    mRNA    linear   PLN 31-JAN-2001
DEFINITION  Oryza sativa subsp. japonica nuclear cap-binding protein CBP80
            mRNA, complete cds.
ACCESSION   AY017415
VERSION     AY017415.1  GI:12621069
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 3009)
  AUTHORS   Derba,M., Kmieciak,M. and Jarmolowski,A.
  TITLE     Molecular characterization of Oryza sativa CBP80
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3009)
  AUTHORS   Derba,M., Kmieciak,M. and Jarmolowski,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2001) Gene Expression, Institute of Molecular
            Biology and Biotechnology, Miedzychodzka 5, Poznan 60-371, Poland
FEATURES             Location/Qualifiers
     source          1. .3009
                     /organism="Oryza sativa (japonica cultivar-group)"
```

```
                     /sub_species="japonica"
                     /db_xref="taxon:39947"
                     /clone="D40981"
                     /dev_stage="shoot"
     CDS             76. .2808
                     /note="similar to Homo sapiens CBP80 and to rice product
                     encoded by GenBank Accession Number D40981"
                     /codon_start=1
                     /product="nuclear cap-binding protein CBP80"
                     /protein_id="AAG54079.1"
                     /db_xref="GI:12621070"
                     /translation="MSAGWRTLLLRIGDRCPEYGGSVDHKEHIETCYGVLCREYEHSK
                     DAMFEFLLQCADQLPHKIPFFGVLIGLINLENEDFSKGIVDTTHANLQDALHNENRDR
                     IRILLRFLCGLMCSKVVLPNSIIETFEALLSSAATILDEETGNPSWQPRADFYVYCIL
                     ASLPWGGSELFEQVPDEFERVLVGIQSYISIRRHFDDIAFSVFETDEGNSPNKKDFIE
                     DLWERIQVLSRNGWKVKSVPKPHLSFEAQLVAGVSHRFSPISCPHLLSRNASSEIVKG
                     QEKHEADLKYPQRLRRLHIFPTNKAENMQPVDRFVVEECILDVLLFFNGCRKECAFYL
                     VSLPVPFRYEYLMAETIFSQLLLLPNPPFRPIYYTLVIIDLCKALPGAFPSVVVGAVH
                     ALFDRISNMDMECRTRLILWFSHHLSNFQFIWPWQEWAYVKDLPKWAPQRVFVQEVLE
                     REIRLSYFDKIKQSIEDAVELEELLPPKAGPNFRYHSDEGKESTDGHRLSKELVAMVR
                     GRKTQGDIISWVDEKIIPVNGAKFALDVVSQTLLDIGSKSFTHLITVLERYGQIISKL
                     CPNEEMQLLLMDEVSAYWKNSTQMIAIAIDRMMGYRLLSNLAIVKWVFSPANVDQFHV
                     SDRPWEILRNAVSKTYNRIFDLRKEIQTLRKGLQAAKEASEKAARELEEAKSIIEIVD
                     GQPVPSENPGRLRRLQARADKAKEGEVTTEESLEAKEALLARGLEESKELLRLLFKSF
                     VEVLTERLPPISADGDVPNLRAGDPNVNSSARDPEATTMEIDNENGGDNDSQLNGQNK
                     KISHNVGELEQWCLCTLGYLKSFSRQYATEIWSHIAMLDQEIFVGNIHPLIRKAASPV
                     CADLPVKGLTFDFDPSSELEFTCIIRDKFFIMLSLTGHAQKYSDVYSGTPSILL"
BASE COUNT      821 a    623 c    683 g    881 t      1 others
ORIGIN
```

```
 Query Match             36.0%;  Score 978.2;  DB 8;  Length 3009;
 Best Local Similarity   62.5%;  Pred. No. 3e-229;
 Matches 1642;  Conservative    0;  Mismatches  929;  Indels   57;  Gaps    5;

Qy      46 tggaaaactcttctccttcgcatcggcgaaaagggacctgagtacggcacttcctccgac 105
           ||||  || || || || || |||||||| | | | || ||||||||     |||  ||||
Db      88 TGGAGGACGCTGCTGCTGCGGATCGGCGACAGGTGCCCGGAGTACGGGGGCTCCGTCGAC 147

Qy     106 tacaaagaccacatcgagacttgtttcggtgtcattcgtagagaaatcgagcgttctgga 165
           |||| || |||||||| |||||||  |||||  || || |||||  ||| |  ||   |
Db     148 CACAAGGAGCACATCGAAACTTGTTATGGTGTGCTTTGTCGAGAATACGAACACTCCAAA 207

Qy     166 gatcaagttttgccttttctactacaatgtgctgaacaattgcctcataagattccttttg 225
           |||  | | ||    ||||| || |||||||| || ||||||||||| ||||||||||
Db     208 GATGCAATGTTTGAGTTTCTCCTCCAATGTGCAGATCAATTGCCTCACAAGATTCCTTTC 267

Qy     226 tatgggactttgattggtttgttgaacttggagaatgaagattttgtccagaagctagta 285
           | |||    ||||| |||||| | |||||||| ||||||||||||  | ||    | ||
Db     268 TTTGGAGTATTGATAGGTTTGATAAACTTGGAAAATGAAGATTTTTCCAAGGGTATTGTC 327

Qy     286 gaaagtgtccacgctaatttccaggtcgctttagattctggcaactgcaacagtatccgt 345
           || |     || || ||||| |||  || ||  ||  ||  ||  |  |||| ||| |
Db     328 GATACAACACATGCCAATTTACAGGATGCCTTGCATAATGAAAATCGTGACAGAATCAGG 387

Qy     346 atattgcttcgctttatgacttccctgttgtgcagtaaggttattcaacctgcttctttg 405
           |||||||| || ||| |   |  |||| ||||||| || ||| | |  ||   |||| |
Db     388 ATATTGCTGCGNTTTCTCTGTGGCCTGATGTGCAGCAAAGTTGTCCTGCCAAATTCTATT 447

Qy     406 attgtcgtcttcgaaacattgctatcatctgctgccactactgtggatgaagagaaagga 465
           ||||     || ||  || |  ||||||||||||| || |   | |||||  || |  |||
Db     448 ATTGAAACATTTGAGGCACTATTATCATCTGCTGCAACAATATTAGATGAGGAAACCGGA 507

Qy     466 aatccatcatggcagccacaagctgacttttacgttatatgcatcttgtccagcctcccg 525
           ||||| || ||||| |||| |||| || || |||   || |||||| |   ||| ||
Db     508 AATCCTTCGTGGCAACCACGTGCTGATTTCTATGTTTATTGTATCTTGGCTTCCCTTCCA 567

Qy     526 tggggaggatcagaactcgctgagcaagttcctgatgagattgaaagagtgttagttggg 585
           ||||| || |||||| |   |||||||||||| |||||  |||| |||||  | |||||
Db     568 TGGGGTGGCTCAGAATTGTTTGAGCAAGTTCCAGATGAATTTGAGAGAGTTCTGGTTGGT 627

Qy     586 atacaagcttatttgagcatccgaaagaattcttccacctctgggttaaacttttttcac 645
           |||||  ||||| | |||||  ||| | ||| |        ||  ||     | ||| |
```

```
Db     628 ATACAGTCTTATATAAGCATTAGAAGGCATTTTGATGATATTGCTTTCTCAGTCTTTGAA 687

Qy     646 aacggagaatttgaaagcagccttgcagagaaggatttcgtggaggatctattggatcga 705
           |  |  |||   |  | |   |       | |||||||| | || ||| ||| ||| ||
Db     688 ACAGATGAA---GGCAACTCTCCCAACAAAAAGGATTTCATCGAAGATTTATGGGAGCGT 744

Qy     706 attcagtctctggcttccaatggatggaaacttgaaagcgtacctaggcctcatctctcg 765
           |||||   |||  ||  |||||| |||||  || | || || || ||  | |||| || |||
Db     745 ATTCAAGTTCTTTCTCGCAATGGGTGGAAGGTTAAGAGTGTTCCAAAACCTCACCTGTCG 804

Qy     766 tttgaagctcaactcgttgctggaaagtttcatgagctacgtcccattaaatgtatggaa 825
           ||||||||||| || || ||||||   | |||      |    || |||| ||      |
Db     805 TTTGAAGCTCAGCTGGTAGCTGGAGTTTCTCACCGTTTCTCCCCAATTAGTTGCCCCCAC 864

Qy     826 caaccgagtccaccttctgatcattcgagggcatacagtggcaagcaaaagcatgatgca 885
           | ||   ||    | |       |   | | |   | ||  | ||||||||||| ||
Db     865 CTACTATCTCGCAATGCATCCTCTGAAATAGTA---AAAGGTCAAGAAAAGCATGAAGCT 921

Qy     886 ttgacgagatatccccagagaattcgtaggttgaatatatttccagctaataaaatggag 945
           ||  |||| || ||  |||||||| |  | |||||||| | ||||||   |||
Db     922 GATCTGAAGTATCCTCAAAGGCTTCGTAGGCTTCACATATTTCCAACAAATAAAGCTGAG 981

Qy     946 gatgtacaaccaattgatcgttttgtcgtggaggagtatttgctggatgtgctcttctat 1005
           |  | |||||  | |||||||||| || || || |   |  ||||||||| | |
Db     982 AACATGCAACCTGTAGATCGTTTTGTTGTTGAAGAATGCATATTGGATGTGCTACTTTTC 1041

Qy    1006 ttgaatggatgtcggaaggagtgtgcatcctacatggctaatcttcctgttacatttcgg 1065
           || |||||||||| || || |||||||  ||  |||  |   | |||||  | || ||
Db    1042 TTCAATGGATGTCGCAAAGAATGTGCATTTTATCTGGTCAGCTTACCTGTGCCTTTCCGT 1101

Qy    1066 tacgagtatcttatggcagagacactattttctcagatactgctgctaccccagccacca 1125
           || || || || ||||| |||||  ||||||| ||| |||||  |  | ||  | || ||
Db    1102 TATGAATACCTGATGGCTGAGACCATATTTTCACAGCTACTGTTATTGCCGAATCCCCCT 1161

Qy    1126 ttcaagactctttattatacactcgtgattatggatctttgtaaggctcttccgggtgcc 1185
           |||| | |  |||| |||||  | || |||| || ||||| |||||  | || |||||
Db    1162 TTCAGGCCAATTTACTATACCTTGGTTATTATCGACCTTTGCAAGGCATTGCCAGGTGCA 1221

Qy    1186 tttcctgctgttgttgctggcgctgttcgtgcactatttgagaaaatatccgacttagac 1245
           |||||| | ||||| |  || || || | ||| || ||||| | |||   | || | ||
Db    1222 TTTCCTTCAGTTGTGGTAGGAGCAGTACATGCTCTTTTTGACAGAATTAGCAACATGGAT 1281

Qy    1246 atggaatccaggacgcgtcttatcctctggttttctcaccacttatccaacttccaattc 1305
           ||||| | | || || |||||||| |||||||| || ||||| || || || || |||
Db    1282 ATGGAGTGCCGCACCCGACTTATCCTATGGTTTTCACATCACTTGTCAAATTTTCAGTTC 1341

Qy    1306 atctggccgtgggaagagtgggcttttgtgttggatcttcccaagtgggcccctaagcgt 1365
           || ||||| ||| | |||||||||  ||  ||| ||||| || ||||| ||  |||||
Db    1342 ATTTGGCCTTGGCAGGAGTGGGCTTACGTCAAGGACCTTCCAAAATGGGCTCCACAGCGT 1401

Qy    1366 gtatttgttcaggagattttgcaaagagaagtacgcttgtcttactgggataaaattaag 1425
           || ||||| || ||  | ||  |||| ||| | |||||||| ||||  || |||||||||
Db    1402 GTTTTTGTCCAAGAAGTATTAGAAAGGGAAATTCGCTTGTCCTACTTTGACAAAATTAAG 1461

Qy    1426 cagagcattgagaatgcgactgccctagaagaattacttcctccaaaagctggtccgaat 1485
           |||||||| ||| ||||  ||  | |||||| |  | || ||||||||||| || ||
Db    1462 CAGAGCATAGAGGATGCTGTTGAATTGGAAGAACTGTTACCCCCAAAAGCTGGGCCTAAC 1521

Qy    1486 tttatgtattccttggaagaaggtaaagagaaaacagaagaacagcaattgtcagccgaa 1545
           || |  |||      || |||| |||| |  || || |  ||  | | ||    |||
Db    1522 TTCAGATATCATAGTGATGAAGGCAAAGAAAGCACTGATGGCCACAGACTCTCCAAGGAA 1581

Qy    1546 ttgagcaggaaggtcaaggaaaaacaaaccgcacgtgacatgatagtgtggattgaagaa 1605
           |     | ||| |  | ||   | ||   | |||| || ||    ||| | || |||
Db    1582 CTTGTTGCCATGGTTAGAGGAAGGAAGACACAAGGTGATATTATTTCATGGGTAGACGAA 1641

Qy    1606 acgatatatccagttcatggttttgaagttactcttacaatagttgtacagaccttactt 1665
           |  |||  ||| ||  |||||    || || | |||    |||||   |||||  | ||
Db    1642 AAAATAATTCCTGTAAATGGTGCCAAATTTGCACTTGATGTAGTTAGCCAGACACTTCTG 1701

Qy    1666 gacatcggatcaaaaagtttcactcatttggtcactgtcctggagcgatatggccaagta 1725
           ||||| || |||||||||||||| ||| |  |||||||  |||||  |||||| ||| ||
Db    1702 GACATTGGCTCAAAAAGTTTCACCCATCTTATCACTGTTTTGGAGAGATATGGTCAAATA 1761
```

```
Qy   1726 ttttcaaagctttgtcctgataacgataagcaggtgatgctattatctcaagtgagtaca 1785
          | ||||||||| || ||  || | || | |||| |  || |  |   | |||| ||| |
Db   1762 ATATCAAAGCTGTGCCCGAATGAAGAAATGCAGTTATTGTTGATGGATGAAGTCAGTGCT 1821

Qy   1786 tactggaaaaacaatgtacaaatgacggcggtggcaattgataggatgatgggttataga 1845
          || ||||| |||| |   || |||  ||  | || ||||||||||||||||||||| |
Db   1822 TATTGGAAGAACAGTACCCAGATGATTGCCATAGCTATTGATAGGATGATGGGTTATCGC 1881

Qy   1846 ctagtatctaatcaggcaattgttagatgggtgttctctccagaaaatgttgatcagttt 1905
          ||| | || |||| ||| || || | |||||| || ||||| |  |||||||||| |||
Db   1882 CTACTTTCCAATCTGGCTATAGTCAAATGGGTTTTTTCTCCTGCTAATGTTGATCAATTT 1941

Qy   1906 catgtgtctgatcagccatgggagatacttggcaatgctcttaacaagacttataaccgt 1965
          ||||| || ||||  |||||||||| ||| | |||||| |||  || || || || ||
Db   1942 CATGTTTCAGATCGTCCATGGGAGATTCTTAGAAATGCTGTTAGTAAAACATACAATCGG 2001

Qy   1966 atctctgatttgaggaaagatatatcaaacattacgaaaaatgttttggttgctgagaaa 2025
          || | |||  |  |||||||| ||    |  | | ||||  | ||   | |||| |  |
Db   2002 ATTTTTGACCTCCGGAAAGAAATTCAGACACTCAGGAAAGGTCTTCAAGCTGCTAAAGAG 2061

Qy   2026 gcttcagccaatgcacgagtagagttggaggctgctgagagcaaactttccctagtggaa 2085
          ||    |  || ||      |||||||||||  ||| |     |   ||    | || ||
Db   2062 GCCAGTGAAAAGGCCGCCAGAGAGTTGGAGGAGGCTAAATCTATTATTGAGATTGTAGAT 2121

Qy   2086 ggtgaacccgttcttggtgagaatccagcgaagatgaagcgtttaaaatcaacagtggag 2145
          ||  |||| || |    ||| |||||||  | | | |  ||  |  || |    |  ||
Db   2122 GGCCAACCTGTGCCATCTGAAAATCCAGGAAGGCTAAGACGACTTCAAGCGCGTGCTGAC 2181

Qy   2146 aagacaggggaagcggagttatctcttcgggagtccctagaggcaaaagaggctcttctt 2205
          ||  |    |||| ||  || | |   || ||  |||| ||||| ||||| || |||
Db   2182 AAAGCGAAAGAAGGAGAAGTAACCACTGAAGAATCTTTAGAAGCAAAGGAGGCCCTCCTT 2241

Qy   2206 aacagagctctctctgagaccgaggttttactgctcttgctgttccaaagtttcttaggt 2265
             |||  ||    || | | |||  || ||    || || ||| | || ||  | |
Db   2242 GCTCGAGGGCTTGAAGAAAGCAAGGAATTGCTTAGGTTACTATTCAAGAGCTTTGTTGAA 2301

Qy   2266 gtactgaaggaacggctcccagatccaactaaagtgagatcagtgcaggatctaaaatct 2325
          || || |  |||||  | |||  |    ||   |   ||   || |   || ||    ||
Db   2302 GTGCTAACTGAACGTTTGCCACCTATTTCTGCTGATGGAGATGTTCCTAATTTACGTGCT 2361

Qy   2326 atag----------------gtgctgaagatgacaagccatctgcgatggacgtggac 2367
            ||                ||    ||      |  || |  | |||||  | |||
Db   2362 GGAGACCCGAATGTAAATTCTTCAGCCCGTGACCCTGAAGCAACAACCATGGAAATAGAC 2421

Qy   2368 agcgagaatgg----------------------------aaacccaaagaagagttgc 2397
          |  || |||||                          ||||   || |  |||  |
Db   2422 AATGAAAATGGAGGAGATAACGATAGTCAGCTGAATGGTCAAAACAAGAAAATCAGTCAC 2481

Qy   2398 gaagtcggtgagagagaacagtggtgcttatcaacacttggctatctcacggcatttaca 2457
            | || || |||   || || |||||  | |  ||| | |||||||||| | | ||| |
Db   2482 AATGTTGGAGAGCTTGAGCAATGGTGTCTCTGCACATTGGGCTATCTCAAGTCGTTTTCT 2541

Qy   2458 aggcaatatgcgagcgagatatggcctcacatggagaagttggagtcagaagtgttctcg 2517
            | |||||||| |  ||||| ||| | || || |  | ||||||    ||  | |||
Db   2542 CGTCAATATGCTACTGAGATCTGGTCCCATATTGCCATGTTGGATCAGGAGATTTTC--- 2598

Qy   2518 ggtgaagatgtgcatcctctctttctccaagccatatcttctgcacttcaattcccatta 2577
          | ||   || | || ||||| . | |   ||||    ||| | |    |  |  || |  |
Db   2599 GTTGGGAATATTCACCCTCTTATCCGGAAAGCTGCTTCTCCGGTTTGTGCAGACCTACCA 2658

Qy   2578 cattaatcttcctctttcaatctcaatcaaacctgtctcttttgtttt 2625
             |   | | ||  || | |  | |   || ||    ||    |||
Db   2659 GTGAAGGGTCTCACCTTTGATTTTGACCCTTCCAGTGAGTTGGAGTTT 2706
```

```
RESULT   7
NCB13O20
LOCUS       NCB13O20               65076 bp    DNA     linear   PLN 13-NOV-2001
DEFINITION  Neurospora crassa DNA linkage group V BAC clone B13O20.
ACCESSION   AL451015
VERSION     AL451015.1  GI:11595712
KEYWORDS    .
```

```
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (sites)
  AUTHORS   Schulte,U., Aign,V., Hoheisel,J., Brandt,P., Fartmann,B.,
            Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 65076)
  AUTHORS   German Neurospora genome,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2001) MIPS, Institut fuer Bioinformatik,
            GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,
            Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:
            G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute
            of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,
            E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT     BAC clone 13O20 (strain OR74A) is available at the Fungal Genetic
            Stock Center, http://www.fgsc.net
            Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
            http://www.mwgdna.com
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of linkage groups II
            and V can be viewed at: http://mips.gsf.de/proj/neurospora.
FEATURES             Location/Qualifiers
     source          1. .65076
                     /organism="Neurospora crassa"
                     /db_xref="taxon:5141"
                     /chromosome="2"
     misc_feature    1. .2829
                     /note="overlap to BAC B7A16, EMBL:AL513445"
     repeat_region   4914. .4950
                     /note="37 bp poly(t) tandem repeat"
     exon            complement(10905. .11174)
                     /gene="B13O20.030"
                     /number=1
     gene            complement(join(10905. .11174,11204. .11227))
                     /gene="B13O20.030"
     CDS             complement(join(10905. .11174,11204. .11227))
                     /gene="B13O20.030"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAC18193.1"
                     /db_xref="GI:11595715"
                     /db_xref="SPTREMBL:Q9HE99"
                     /translation="MALNWRLSRYLKKRIPSAMGAPLRSFWIFESRCVHAALWQAPAE
                     ASFCGGTSLAGWLAALDWLAKLRLLGLEAVVNLTPDRRHKLNWKSVRVGGGVH"
     gene            10905. .11227
                     /gene="B13O20.030"
     intron          complement(11175. .11203)
                     /gene="B13O20.030"
                     /number=1
     exon            complement(11204. .11227)
                     /gene="B13O20.030"
                     /number=2
     gene            15224. .17874
                     /gene="B13O20.050"
     CDS             join(15224. .15248,15603. .15643,15668. .15719,15962. .15974,
                     16168. .16822,16871. .17080,17136. .17309,17374. .17874)
                     /gene="B13O20.050"
                     /note="similarity to finger protein AZF1, asparagine-rich,
                     Saccharomyces cerevisiae, PIR:S46593
                     Contains Zinc finger, C2H2 type, domain [CQIEGCN
                     THKRRH][CEVCDKGFVQRGDVKAHMKTH][CRLDNCHKQFTQRGNLKYHQNKYH]"
                     /codon_start=1
                     /product="related to finger protein AZF1"
                     /protein_id="CAC18195.2"
                     /db_xref="GI:16944684"
                     /db_xref="SPTREMBL:Q9HE97"
                     /translation="MHTYTDMYTFAHRLHTPRSVQSASTCRARLLSRLPPIWRGTGPR
                     YPINRMALTAQPSAPAGWERWPQHHPSNDYVMMDADVMPYNARPMTTAPMQQRPSLVP
                     HYMPTTSMSTASINSISAPAHYQSPVPYGGGYPAYPLPTPTTMNSSYHKHQQYQERPS
                     LHMITPEFEDTRGPHHIRNARRYSEESRSPSERSDSQGSTTETTISNHSSCSRTITPN
```

```
                TPVNGAPQVEFSTAVDKLMKVIQSKMKDADPEQSGDDKDIKAEQQSSPVCQARVQQPA
                DKHKRKRYECQIEGCNKKFSQKTHRDTHVRSHTGDRPYVCPIPGCGGRFTQAGNLKTH
                KRRHTGERPYRCEVCDKGFVQRGDVKAHMKTHLGTKAFLCRLDNCHKQFTQRGNLKYH
                QNKYHNETIKALAARFDAIEDWSTVLKEDMEIFKDFAEVHKNSNKGIKGRGKHRKVKS
                VPLSSPTSPTGHSPLPNIMASQYPLPSAPGVSQLLHTPHPLHPPLHHQGPSHSAMYGM
                PRSSLHSHYETYDHHEVDTVASSRGSVAEPIYHHEEHPRELAFGDRMY"
exon            15224. .15248
                /gene="B13O20.050"
                /number=1
intron          15249. .15602
                /gene="B13O20.050"
                /number=1
exon            15603. .15643
                /gene="B13O20.050"
                /number=2
intron          15644. .15667
                /gene="B13O20.050"
                /number=2
exon            15668. .15719
                /gene="B13O20.050"
                /number=3
intron          15720. .15961
                /gene="B13O20.050"
                /number=3
exon            15962. .15974
                /gene="B13O20.050"
                /number=4
intron          15975. .16167
                /gene="B13O20.050"
                /number=4
exon            16168. .16822
                /gene="B13O20.050"
                /number=5
intron          16823. .16870
                /gene="B13O20.050"
                /number=5
exon            16871. .17080
                /gene="B13O20.050"
                /number=6
intron          17081. .17135
                /gene="B13O20.050"
                /number=6
exon            17136. .17309
                /gene="B13O20.050"
                /number=7
intron          17310. .17373
                /gene="B13O20.050"
                /number=7
exon            17374. .17874
                /gene="B13O20.050"
                /number=8
tRNA            19101. .19172
                /note="tRNA predict as a tRNA- Cys : anticodon gca"
                /product="tRNA-Cys"
gene            complement(join(20030. .20961,21020. .21143))
                /gene="B13O20.060"
gene            20030. .21143
                /gene="B13O20.060"
CDS             complement(join(20030. .20961,21020. .21143))
                /gene="B13O20.060"
                /note="similarity to carrier protein FLX1, inner membrane,
                Saccharomyces cerevisiae, PIR:S48400
                Contains Mitochondrial energy transfer proteins signature"
                /codon_start=1
                /product="related to FAD carrier protein FLX1"
                /protein_id="CAC18196.1"
                /db_xref="GI:11595718"
                /db_xref="SPTREMBL:Q9HE96"
                /translation="MTKSSDAGLSPALVETVAGLSAGSMATLIVHPLDIVKTRMQVHR
                SSPTNPSAALTTVSVFRSLAQTDRPLAALYRGLTPNLIGNATSWASFFFFKSRFERLI
                AHLKAPSSPPAPSHLGGEPTTLPQNVTPDRETQAQIKSHLSPTDFFAASLLAGAATQI
                ITNPIWVLKTRMLSTDRLAADAYPSMFTGAVRLFRSEGILGFYRGLGVGMLAISHGAV
                QFAVYDPARRMYIASRDAKRRLAGQEIASEDERESQRISNEATIVLSTAAKLVAGTAT
                YPLQVMRARLQHHQADELFGRGIGGVVAKLWREEGFRGFYRGMMPGVVRVLPATWVTF
                LVYENVKYYLPKMVSGV"
exon            complement(20030. .20961)
                /gene="B13O20.060"
                /number=1
intron          complement(20962. .21019)
                /gene="B13O20.060"
                /number=1
exon            complement(21020. .21143)
                /gene="B13O20.060"
                /number=2
gene            23128. .24376
                /gene="B13O20.070"
exon            23128. .23141
                /gene="B13O20.070"
                /number=1
CDS             join(23128. .23141,23218. .24376)
                /gene="B13O20.070"
                /note="strong similarity to spliceosome-associated protein
                SAP-49, Homo sapiens, PIR:A54964
                Contains Eukaryotic putative RNA-binding region RNP-1
                signature [KGYGFVSF]"
                /codon_start=1
                /product="probable spliceosome-associated protein SAP-49"
                /protein_id="CAC18197.1"
                /db_xref="GI:11595719"
                /db_xref="SPTREMBL:Q9HE95"
                /translation="MSKHWEQNKDATVYVGNIDERFTQELLSELMTQVGPVRQVHMPQ
                DRVSQTHQGYGFVEFDTPASAEYAAKVLNGIRVWGKPIRVNKASADKQKTVDIGAELF
                INNLDPQVDEKILYDTFSQFGQILRQPNIVRDDNNISKGYGFVSFGSFEASDAARATM
                NGQYLLSKQITVEYAYKKDGKGERHGDEAERKLAAEGKKHNIVPEQQPLPPAFHMTRP
                PAPGIAPAADIPTAAVIPPAAVAPAVPIPVGVPVAAAPAYGVPPPGPAAMGRSGPPYG
                AGMPPAGGRIPNLPPAPSGLPARPPPSHGGFGGPVGDFHPGGTAFRGGPPGPPGPAVP
                PGPPAIYPGMAPPAGFPGAPPLGASMPPPGFMPPPGAAPPPGFGAPPGAPNGFARR"
intron          23142. .23217
                /gene="B13O20.070"
                /number=1
exon            23218. .24376
                /gene="B13O20.070"
                /number=2
repeat_region   26292. .26324
                /note="33 bp gcagcagcagcagcagcagca tandem repeat"
repeat_region   26390. .26481
                /note="92 bp ttttttttttttttttttttttttttttttttttttttttttt
                ttttttttttttt. . . tandem repeat"
gene            33747. .34190
                /gene="B13O20.090"
gene            complement(33747. .34190)
                /gene="B13O20.090"
```

```
  Query Match             2.5%;  Score 69.2;  DB 8;  Length 65076;
  Best Local Similarity   52.4%;  Pred. No. 3.1e-06;
  Matches  152;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy   1069 gagtatcttatggcagagacactattttctcagatactgctgctaccccagccaccattc 1128
          ||| || |    |  ||  |  | || || ||  |  |   | | ||| | ||      |
Db  40704 GAGGATGTCGCTGTTGACGCGGTCTTCTCGCAACTCTTCTCGTTGCCCAATCCCGAGCAC 40763

Qy   1129 aagactctttattatacactcgtgattatggatctttgtaaggctcttccgggtgccttt 1188
          |||    |||||||      |||  ||| |||   ||| |||      |||| |||| |
Db  40764 AAGCTCGTTTATTACCATGCCGTCCTTACGGAGGCTTGCAAGCTGGCACCGGCTGCCATC 40823

Qy   1189 cctgctgttgttgctggcgctgttcgtgcactatttgagaaaatatccgacttagacatg 1248
           || |    | | | | ||| |||||   || |    ||| |    |  | | ||| ||
Db  40824 GCTCCCAGCTTAGGTCGTGCTATTCGTTTCCTTTACAGGAATAGCCACCGCATGGACTTG 40883

Qy   1249 gaatccaggacgcgtcttatcctctggttttctcaccacttatccaacttccaattcatc 1308
          ||       |  ||  | ||    ||||| || ||||| ||    || |||  ||| | |
Db  40884 GAGCTTTCCAATCGGTTCATGGATTGGTTCTCACACCATTTGAGTAATTTCGGATTTACC 40943

Qy   1309 tggccgtgggaagagtgggcttttgtgttggatcttcccaagtgggcccc 1358
          |||  ||||   |||||||         |  |||| |||| | |. |||
Db  40944 TGGAAGTGGACCGAGTGGGTGGACAATGTTTATCTGTCCAACTTGCACCC 40993
```

```
RESULT   8
DME238970
LOCUS       DME238970               3228 bp    mRNA    linear   INV 19-AUG-1999
DEFINITION  Drosophila melanogaster mRNA for cap binding protein 80 (Cbp80).
ACCESSION   AJ238970
VERSION     AJ238970.1  GI:5763574
KEYWORDS    cap binding protein 80; cbp80 gene.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 3228)
  AUTHORS   Lewis,J.D.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3228)
  AUTHORS   Lewis,J.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1999) Lewis J.D., Institute of Cell and Molecular
            Biology, University of Edinburgh, Kings Buildings, Mayfield Road,
            Edinburgh, Lothian, EH9 3JR, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .3228
                     /organism="Drosophila melanogaster"
                     /strain="Oregon R"
                     /db_xref="taxon:7227"
                     /map="90E1"
                     /tissue_type="imaginal disc"
                     /dev_stage="third instar"
     gene            269. .2671
                     /gene="Cbp80"
     CDS             269. .2671
                     /gene="Cbp80"
                     /function="putative role in pre-mRNA splicing"
                     /codon_start=1
                     /product="cap binding protein 80"
                     /protein_id="CAB53186.1"
                     /db_xref="GI:5763575"
                     /translation="MSRRRAHDTEDEGYDHRRNKRRRVSENQEIEDRLESLILRVGEA
                     TTSSVESNLEGLVSVLEADLGTFRLKILRILSDCAVAMPEKCTVYTTLVGLLNAKNYK
                     FGGEFVDHMVKTFKESLKMCRWDAARYSLRFLADLVNCHVISATSLLQLLDTMIDVSN
                     EDTVPQVRRDWFVFAVLSTLPWVGRDLYEKKESALESLLLRIEVYLNKRSKKHHNALR
                     VWSSDAPHPQEEYLDCLWAQIRKLRQDNWAEKHIPRPYLVFDSILCEALQHNLPTIVP
                     PPHHDNFEYPMPWVVYRMFDYTDCPDGPNLPGAHSIERFLIEEHLHHIIETYHHERKD
                     CAAQLLSFPYKHKIPLEYCIVEVVFAELFHMPTPRYLDICYGSILIELCKLQPATLPQ
                     VLAQATEILFMRIDSMNTSCFDRFVNWFSYHLSNFKFTWSWDEWDSCLLLDGEHPRPK
                     FIQEVLQKCLRLSYHQRITEMMPTTYAKLIPLTPVPNYKYANEEAANLPGTTVAHQLV
                     VAIRQKCTPEEVVNILKDIPNSGYSGEEMSDGSFNALKIDVFVQTLLNLGSKSFSHSF
                     AAISKFHSVFRALAETEEAQICILHNIFELWSSHQQMMVVLIDKLLKLQIVDCSAVAT
                     WIFSKEMTGEFTKLYLWEILHLTIKKMNKHVIKLNTELSEAKEKLAKADSSSSDSEDD
                     SSHKRKKPITHADKPSEEVVERMEEKLEAANVNQKRLFLIVFQRFIMILSEHLLRSDT
                     DGRDPDTDWYRWTIGRLQQVFLMHHEQVQKYSSTLETLLFTSDLDTHILEVFQQFVAL
                     RA"
BASE COUNT      845 a    807 c    807 g    769 t
ORIGIN
```

```
  Query Match             2.4%;  Score 65;  DB 3;  Length 3228;
  Best Local Similarity   47.1%;  Pred. No. 3.4e-05;
  Matches  237;  Conservative    0;  Mismatches  260;  Indels    6;  Gaps    1;

Qy     937 aaaatggaggatgtacaaccaattgatcgttttgtcgtggaggagtatttgctggatgtg 996
           ||  ||  || ||  ||  |||| || || ||| |  | |||||| || | |   || |
Db    1181 AATCTGCCGGGTGCCCACTCAATAGAGCGCTTTTTGATTGAGGAGCATCTACACCATATC 1240

Qy     997 ctcttctatttgaatggatgtcggaaggagtgtgcatcctacatggctaatcttcc---- 1052
             |       |   ||      || ||||| || ||  || |  ||  ||   ||||
Db    1241 ATTGAAACATACCATCACGAGCGAAAGGATTGCGCCGCCCAGCTGCTTAGCTTTCCGTAC 1300

Qy    1053 --tgttacatttcggtacgagtatcttatggcagagacactattttctcagatactgctg 1110
               |   | | |   ||||||   || |  |||    | ||  |  || |  | |
```

```
Db    1301 AAGCACAAGATCCCGCTGGAGTATTGCATCGTCGAGGTGGTCTTCGCCGAGCTCTTTCAT 1360

Qy    1111 ctaccccagccaccattcaagactctttattatacactcgtgattatggatctttgtaag 1170
            | |||  ||| |  |    |   | | |||      | |  | || || || || ||
Db    1361 ATGCCCACGCCGCGCTATCTGGACATCTGTTACGGCTCCATTTTGATCGAACTCTGCAAA 1420

Qy    1171 gctcttccgggtgcctttcctgctgttgttgctggcgctgttcgtgcactatttgagaaa 1230
              |  ||||   |  | ||    ||| | ||    ||          || |||  |
Db    1421 CTGCAGCCGGCCACGCTGCCCCAGGTTCTGGCCCAGGCCACCGAGATCCTCTTTATGCGC 1480

Qy    1231 atatccgacttagacatggaatccaggacgcgtcttatcctctggtttttctcaccactta 1290
           ||   |    |  ||| |  ||        ||  |  ||   ||||| ||  | ||  |
Db    1481 ATCGACTCGATGAACACGTCATGTTTCGATCGGTTCGTCAATTGGTTCTCCTATCATCTG 1540

Qy    1291 tccaacttccaattcatctggccgtgggaagagtgggcttttgtgttggatcttcccaag 1350
             ||||||| | ||||  ||| | ||||| || ||||        ||   ||   |
Db    1541 AGCAACTTCAAGTTCACATGGTCATGGGACGAATGGGACAGCTGTCTGCTCCTCGACGGC 1600

Qy    1351 tgggcccctaagcgtgtatttgttcaggagattttgcaaagagaagtacgcttgtcttac 1410
             |   ||    |   |||  | |||||| | ||||| |      ||||| | || ||
Db    1601 GAGCATCCGCGTCCCAAATTCATCCAGGAGGTGTTGCAGAAGTGCCTACGCCTTTCGTAT 1660

Qy    1411 tgggataaaattaagcagagcat 1433
             |    |||| | | |  ||
Db    1661 CATCAGCGCATTACGGAAATGAT 1683
```

```
RESULT   9
AY051796
LOCUS       AY051796                2975 bp    mRNA    linear   INV 27-AUG-2001
DEFINITION  Drosophila melanogaster LD31211 full length cDNA.
ACCESSION   AY051796
VERSION     AY051796.1  GI:15291902
KEYWORDS    FLI_CDNA.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2975)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
            Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
            Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K.,
            Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2001) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
            This clone was sequenced as part of a high-throughput process to
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000). The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone.  However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .2975
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
```

```
                     /map="4C7-4C7"
     gene            1. .2975
                     /gene="Cbp80"
                     /note="alignment with genomic scaffold AE003431"
                     /db_xref="FLYBASE:FBgn0022942"
     CDS             256. .2658
                     /gene="Cbp80"
                     /note="Longest ORF"
                     /codon_start=1
                     /db_xref="FLYBASE:FBgn0022942"
                     /product="LD31211p"
                     /protein_id="AAK93220.1"
                     /db_xref="GI:15291903"
                     /translation="MSRRRAHDTEDEGYDHRRNKRRRVSENQEIEDRLESLILRVGER
                     STSSVESNLEGLVSVLEADLGTFRLKILRILSDCAVRMPEKCTVYTTLVGLLNAKNYK
                     FGGEFVDHMVKTFKESLKMCRWDAARYSLRFLADLVNCHVISATSLLQLLDTMIDVSN
                     EDTVPQVRRDWFVFAVLSTLPWVGRDLYEKKESALESLLLRIEVYLNKRSKKHHNALR
                     VWSSDAPHPQEEYLDCLWAQIRKLRQDNWAEKHIPRPYLVFDSILCEALQHNLPTIVP
                     PPHHDNFEYPMPWVVYRMFDYTDCPDGPNLPGAHSIERFLIEEHLHHIIETYHHERKD
                     CAAQLLSFPYKHKIPLEYCIVEVVFAELFHMPTPRYLDICYGSILIELCKLQPATLPQ
                     VLAQATEILFMRIDSMNTSCFDRFVNWFSYHLSNFKFTWSWDEWDSCLLLDGEHPRPK
                     FIQEVLQKCLRLSYHQRITEMMPTTYAKLIPLTPVPNYKYANEEAANLPGTTVAHQLV
                     VAIRQKCTPEEVVNILKDIPNSGYSGEEMSDGSFNALKIDVFVQTLLNLGSKSFSHSF
                     AAISKFHSVFRALAETEEAQICILHNIFELWSSHQQMMVVLIDKLLKLQIVDCSAVAT
                     WIFSKEMTGEFTKLYLWEILHLTIKKMNKHVIKLNTELSEAKEKLAKADSSSSDSEDD
                     SSHKRKKPITHADKPSEEVVERMEEKLEAANVNQKRLFLIVFQRFIMILSEHLLRSDT
                     DGRDPDTDWYRWTIGRLQQVFLMHHEQVQKYSSTLETLLFTSDLDTHILEVFQQFVAL
                     RA"
BASE COUNT      765 a    772 c    768 g    670 t
ORIGIN
```

```
  Query Match             2.3%;  Score 61.8;  DB 3;  Length 2975;
  Best Local Similarity   46.4%;  Pred. No. 0.00021;
  Matches  241;  Conservative    0;  Mismatches  272;  Indels    6;  Gaps    1;

Qy    937 aaaatggaggatgtacaaccaattgatcgttttgtcgtggaggagtatttgctggatgtg 996
          ||  ||  || ||  ||  |||| || || ||| |  | |||||| || | |   || |
Db   1168 AATCTGCCGGGTGCCCACTCAATAGAGCGCTTTTTGATTGAGGAGCATCTACACCATATC 1227

Qy    997 ctcttctatttgaatggatgtcggaaggagtgtgcatcctacatggctaatcttcc---- 1052
          |       |   ||      || ||||| || ||  || |  ||  ||   ||||
Db   1228 ATTGAAACATACCATCACGAGCGAAAGGATTGCGCCGCCCAGCTGCTTAGCTTTCCGTAC 1287

Qy   1053 --tgttacatttcggtacgagtatcttatggcagagacactatttttcagatactgctg 1110
            |   | | |   |||||    || |  |||    | ||  |  || | | |
Db   1288 AAGCACAAGATCCCGCTGGAGTACTGCATCGTCGAGGTGGTCTTCGCCGAGCTCTTTCAT 1347

Qy   1111 ctaccccagccaccattcaagactctttattatacactcgtgattatggatctttgtaag 1170
           | |||  ||| |  |     |    | |  ||      | |  | || || || || ||
Db   1348 ATGCCCACGCCGCGCTATCTGGACATCTGCTACGGCTCCATTTTGATCGAACTCTGCAAA 1407

Qy   1171 gctcttccgggtgcctttcctgctgttgttgctggcgctgttcgtgcactatttgagaaa 1230
           |  ||||   |  | ||   ||| | ||    ||          || |||  |
Db   1408 CTGCAGCCGGCCACGCTGCCCCAGGTTCTGGCCCAGGCCACCGAGATCCTCTTTATGCGC 1467

Qy   1231 atatccgacttagacatggaatccaggacgcgtcttatcctctggtttttctcaccactta 1290
          ||   |    | ||| |  ||        ||  |  ||   ||||| ||  | ||  |
Db   1468 ATCGACTCGATGAACACGTCATGTTTCGATCGGTTCGTCAATTGGTTCTCCTATCATCTG 1527

Qy   1291 tccaacttccaattcatctggccgtgggaagagtgggcttttgtgttggatcttcccaag 1350
          |||||||  | ||||  ||| | ||||| || ||||          ||   ||   |
Db   1528 AGCAACTTCAAGTTCACATGGTCATGGGACGAATGGGACAGCTGTCTGCTCCTCGACGGC 1587

Qy   1351 tgggcccctaagcgtgtatttgttcaggagattttgcaaagagaagtacgcttgtcttac 1410
           |   ||    |   |||  | |||||| | ||||| |       ||||| | || ||
Db   1588 GAGCATCCGCGTCCCAAATTCATCCAGGAGGTGTTGCAGAAGTGCCTACGCCTTTCGTAT 1647

Qy   1411 tgggataaaattaagcagagcattgagaatgcgactgcc 1449
             |   |||| | | |  ||    |   |   ||||
Db   1648 CATCAGCGCATTACGGAAATGATGCCAACGACTTATGCC 1686
```

```
RESULT  10
HUMNCBP
LOCUS       HUMNCBP                 3066 bp    mRNA    linear   PRI 07-FEB-1999
DEFINITION  Human mRNA for nuclear cap binding protein, complete cds.
ACCESSION   D32002
VERSION     D32002.1  GI:1865647
KEYWORDS    nuclear cap binding protein.
SOURCE      Homo sapiens cell_line:HeLa cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3066)
  AUTHORS   Kataoka,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-1994) Naoyuki Kataoka, Faculty of Science, Kyoto
            University, Department of Biophysics; Kitashirakawa-Oiwake,
            Sakyo-ku, Kyoto 606, Japan
            (E-mail:g53376@sakura.kudpc.kyoto-u.ac.jp, Tel:075-753-4232,
            Fax:075-753-4235)
REFERENCE   2  (bases 1 to 3066)
  AUTHORS   Kataoka,N., Ohno,M., Kangawa,K., Tokoro,Y. and Shimura,Y.
  TITLE     Cloning of a complementary DNA encoding an 80 kilodalton nuclear
            cap binding protein
  JOURNAL   Nucleic Acids Res. 22 (19), 3861-3865 (1994)
  MEDLINE   95023141
COMMENT     On Mar 5, 1997 this sequence version replaced gi:577627.
FEATURES             Location/Qualifiers
     source          1. .3066
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HeLa"
     CDS             265. .2637
                     /codon_start=1
                     /product="nuclear cap binding protein"
                     /protein_id="BAA06769.1"
                     /db_xref="GI:577628"
                     /translation="MSRRRHSDENDGGQPHKRRKTSDANETEDHLESLICKVGEKSAC
                     SLESNLEGLAGVLEADLPNYKSKILRLLCTVARLLPEKLTIYTTLVGLLNARNYNFGG
                     EFVEAMIRQLKESLKANNYNEAVYLVRFLSDLVNCHVIAAPSMVAMFENFVSVTQEED
                     VPQVRRDWYVYAFLSSLPWVGKELYEKKDAEMDRIFANTESYLKRRQKTHVPMLQVWT
                     ADKPHPQEEYLDCLWAQIQKLKKDRWQERHILRPYLAFDSILCEALQHNLPPFTPPPH
                     TEDSVYPMPRVIFRMFDYTDDPEGPVMPGSHSVERFVIEENLHCIIKSHWKERKTCAA
                     QLVSYPGKNKIPLNYHIVEVIFAELFQLPAPPHIDVMYTTLLIELCKLQPGSLPQVLA
                     QATEMLYMRLDTMNTTCVDRFINWFSHHLSNFQFRWSWEDWSDCLSQDPESPKPKFVR
                     EVLEKCMRLSYHQRILDIVPPTFSALCPANPTCIYKYGDESSNSLPGHSVALCLAVAF
                     KSKATNDEIFSILKDVPNPNQDDDDDEGFSFNPLKIEVFVQTLLHLAAKSFSHSFSAL
                     AKFHEVFKTLAESDEGKLHVLRVMFEVWRNHPQMIAVLVDKMIRTQIVDCAAVANWIF
                     SSELSRDFTRLFVWEILHSTIRKMNKHVLKIQKELEEAKEKLARQHKRRSDDDDRSSD
                     RKDGVLEEQIERLQEKVESAQSEQKNLFLVIFQRFIMILTEHLVRCETDGTSVLTPWY
                     KNCIERLQQIFLQHHQIIQQYMVTLENLLFTAELDPHILAVFQQFCALQA"
     polyA_signal    3043. .3048
     polyA_site      3066
BASE COUNT      870 a    660 c    699 g    837 t
ORIGIN
```

```
  Query Match             2.1%;  Score 57.2;  DB 9;  Length 3066;
  Best Local Similarity   51.1%;  Pred. No. 0.0028;
  Matches  134;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy   1644 aatagttgtacagaccttacttgacatcggatcaaaaagtttcactcatttggtcactgt 1703
          | |  ||||||||||  | ||  || | | | | |||   ||||  || |   ||| ||
Db   1899 AGTCTTTGTACAGACTCTGCTACACTTGGCAGCCAAATCATTCAGCCACTCCTTCAGTGC 1958

Qy   1704 cctggagcgatatggccaagtattttcaaagctttgtcctgataacgataagcaggtgat 1763
           || |     | |  | |||| ||   ||  ||   |     | | ||     || |
Db   1959 TCTTGCAAAGTTTCACGAAGTCTTCAAAACCCTAGCTGAAAGTGATGAAGGAAAGTTACA 2018

Qy   1764 gctattatctcaagtgagtacatactggaaaaacaatgtacaaatgacggcggtggcaat 1823
          |  ||       ||  |     |||||  ||| ||  ||| |||| || ||   | |
Db   2019 TGTGCTAAGAGTTATGTTTGAGGTCTGGAGGAACCATCCACAGATGATTGCTGTACTAGT 2078

Qy   1824 tgataggatgatgggttatagactagtatctaatcaggcaattgttagatgggtgttctc 1883
```

```
Db   2079 GGATAAGATGATTCGTACACAAATAGTTGATTGTGCTGCCGTAGCAAATTGGATCTTCTC 2138

Qy   1884 tccagaaaatgttgatcagttt 1905
Db   2139 TTCAGAACTATCTCGTGACTTT 2160


RESULT  11
AC013864
LOCUS       AC013864               14831 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC013864
VERSION     AC013864.1  GI:6437471
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 14831)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210552 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .14831
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     3629 a   3627 c   3766 g   3809 t
ORIGIN


  Query Match             2.1%;  Score 56.8;  DB 2;  Length 14831;
  Best Local Similarity   46.6%;  Pred. No. 0.0035;
  Matches  221;  Conservative    0;  Mismatches  247;  Indels    6;  Gaps    1;

Qy    937 aaaatggaggatgtacaaccaattgatcgttttgtcgtggaggagtatttgctggatgtg 996
Db  12178 AATCTGCCGGGTGCCCACTCAATAGAGCGCTTTTTGATTGAGGAGCATCTACACCATATC 12237

Qy    997 ctcttctatttgaatggatgtcggaaggagtgtgcatcctacatggctaatcttcc---- 1052
Db  12238 ATTGAAACATACCATCACGAGCGAAAGGATTGCGCCGCCCAGCTGCTTAGCTTTCCGTAC 12297

Qy   1053 --tgttacatttcggtacgagtatcttatggcagagacactattttctcagatactgctg 1110
Db  12298 AAGCACAAGATCCCGCTGGAGTACTGCATCGTCGAGGTGGTCTTCGCCGAGCTCTTTCAT 12357

Qy   1111 ctaccccagccaccattcaagactctttattatacactcgtgattatggatctttgtaag 1170
Db  12358 ATGCCCACGCCGCGCTATCTGGACATCTGCTACGGCTCCATTTTGATCGAACTCTGCAAA 12417

Qy   1171 gctcttccgggtgccttttcctgctgttgttgctggcgctgttcgtgcactatttgagaaa 1230
Db  12418 CTGCAGCCGGCCACGCTGCCCCAGGTTCTGGCCCAGGCCACCGAGATCCTCTTTATGCGC 12477

Qy   1231 atatccgacttagacatggaatccaggacgcgtcttatcctctggttttctcaccactta 1290
Db  12478 ATCGACTCGATGAACACGTCATGTTTCGATCGGTTCGTCAATTGGTTCTCCTATCATCTG 12537

Qy   1291 tccaacttccaattcatctggccgtgggaagagtgggcttttgtgttggatcttcccaag 1350
Db  12538 AGCAACTTCAAGTTCACATGGTCATGGGACGAATGGGACAGCTGTCTGCTCCTCGACGGC 12597

Qy   1351 tgggcccctaagcgtgtatttgttcaggagattttgcaaagagaagtacgcttg 1404

Db  12598 GAGCATCCGCGTCCCAAATTCATCCAGGAGGTGTTGCAGAAGTGCCTACGGTAG 12651


RESULT  12
AC023715
LOCUS       AC023715               160622 bp    DNA     linear   INV 01-JAN-2002
DEFINITION  Drosophila melanogaster X BAC RP98-17H6 (Roswell Park Cancer
            Institute Drosophila BAC Library) complete sequence.
ACCESSION   AC023715
VERSION     AC023715.3  GI:18030098
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 160622)
  AUTHORS   Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
            Gocayne,J.D., Tabor,P., Williamson,A., Homsi,F.H.,
            Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
            Ayele,M.A., Scott,G.S., Worley,K.W., Amamatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
            Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
            Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
            Draper,H., Emery-Cohen,A., Ferriera,S., Garg,N.D.S., Houck,J.,
            Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
            Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
            Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
            Scheeler,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,
            Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,
            Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160622)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 160622)
  AUTHORS   Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
```

Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE Direct Submission

JOURNAL Submitted (01-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Jan 1, 2002 this sequence version replaced gi:6997285.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.

FEATURES Location/Qualifiers
source 1. .160622
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="X"
/clone="RP98-17H6"

BASE COUNT 44612 a 34756 c 34800 g 46454 t

ORIGIN

Query Match 2.1%; Score 56.8; DB 3; Length 160622;
Best Local Similarity 46.6%; Pred. No. 0.0034;
Matches 221; Conservative 0; Mismatches 247; Indels 6; Gaps 1;

```
Qy     937 aaaatggaggatgtacaaccaattgatcgttttgtcgtggaggagtatttgctggatgtg 996
           ||  ||  || ||  ||  |||| || || ||| |  | |||||| || | |   || |
Db   21162 AATCTGCCGGGTGCCCACTCAATAGAGCGCTTTTTGATTGAGGAGCATCTACACCATATC 21221

Qy     997 ctctttctatttgaatggatgtcggaaggagtgtgcatcctacatggctaatcttcc---- 1052
           |       |  ||       || ||||| || ||  || |  ||  ||   ||||
Db   21222 ATTGAAACATACCATCACGAGCGAAAGGATTGCGCCGCCCAGCTGCTTAGCTTTCCGTAC 21281

Qy    1053 --tgttacatttcggtacgagtatcttatggcagagacactattttctcagatactgctg 1110
```

```
                |   | | |   |||||    || |  |||    | ||  |  || |  | |
Db   21282 AAGCACAAGATCCCGCTGGAGTACTGCATCGTCGAGGTGGTCTTCGCCGAGCTCTTTCAT 21341

Qy    1111 ctaccccagccaccattcaagactctttattatacactcgtgattatggatctttgtaag 1170
           | |||  ||| |  |     |    | |  ||       | |  | || || || || ||
Db   21342 ATGCCCACGCCGCGCTATCTGGACATCTGCTACGGCTCCATTTTGATCGAACTCTGCAAA 21401

Qy    1171 gctcttccgggtgcctttcctgctgttgttgctggcgctgttcgtgcactatttgagaaa 1230
           |  ||||   |  | ||    ||| | ||    ||          || |||  |
Db   21402 CTGCAGCCGGCCACGCTGCCCCAGGTTCTGGCCCAGGCCACCGAGATCCTCTTTATGCGC 21461

Qy    1231 atatccgacttagacatggaatccaggacgcgtcttatcctctggttttctcaccactta 1290
           ||   |    |  ||| |  ||       ||  |  ||   ||||| ||  | ||  |
Db   21462 ATCGACTCGATGAACACGTCATGTTTCGATCGGTTCGTCAATTGGTTCTCCTATCATCTG 21521

Qy    1291 tccaacttccaattcatctggccgtgggaagagtgggctttttgtgttggatcttcccaag 1350
           |||||||  | ||||  ||| | ||||| || ||||       ||   ||   |
Db   21522 AGCAACTTCAAGTTCACATGGTCATGGGACGAATGGGACAGCTGTCTGCTCCTCGACGGC 21581

Qy    1351 tgggcccctaagcgtgtatttgttcaggagattttgcaaagagaagtacgcttg 1404
           |   ||    |    |||  | |||||| | ||||| |      ||||  | |
Db   21582 GAGCATCCGCGTCCCAAATTCATCCAGGAGGTGTTGCAGAAGTGCCTACGGTAG 21635
```

RESULT 13
AC023747/c
LOCUS AC023747 167089 bp DNA linear INV 28-DEC-2001
DEFINITION Drosophila melanogaster X BAC RP98-38L1 (Roswell Park Cancer Institute Drosophila BAC Library) complete sequence.
ACCESSION AC023747
VERSION AC023747.3 GI:17985875
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 167089)
AUTHORS Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A., Gocayne,J.D., Tabor,P., Williamson,A., Homsi,F.H., Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C., Ayele,M.A., Scott,G.S., Worley,K.W., Amamatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C., Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K., Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E., Draper,H., Emery-Cohen,A., Ferriera,S., Garg,N.D.S., Houck,J., Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C., Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B., Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V., Scheeler,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M., Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O., Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.
TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 167089)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 167089)
AUTHORS Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE Direct Submission

JOURNAL Submitted (28-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Dec 28, 2001 this sequence version replaced gi:6996833.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.

FEATURES Location/Qualifiers
source 1. .167089
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="X"
/clone="RP98-38L1"

BASE COUNT 46421 a 37080 c 37190 g 46398 t
ORIGIN

Query Match 2.1%; Score 56.8; DB 3; Length 167089;
Best Local Similarity 46.6%; Pred. No. 0.0034;
Matches 221; Conservative 0; Mismatches 247; Indels 6; Gaps 1;

```
Qy     937 aaaatggaggatgtacaaccaattgatcgttttgtcgtggaggagtatttgctggatgtg 996
           ||  ||  || ||  ||  |||| || || ||| |  | ||||||  || | |  || |
Db   11871 AATCTGCCGGGTGCCCACTCAATAGAGCGCTTTTTGATTGAGGAGCATCTACACCATATC 11812

Qy     997 ctctttctatttgaatggatgtcggaaggagtgtgcatcctacatggctaatcttcc---- 1052
           |       |   ||      || ||||| || ||  || |  ||  ||   ||||
Db   11811 ATTGAAACATACCATCACGAGCGAAAGGATTGCGCCGCCCAGCTGCTTAGCTTTCCGTAC 11752

Qy    1053 --tgttacatttcggtacgagtatcttatggcagagacactattttctcagatactgctg 1110
             |   | | |   |||||    || |  |||    | ||  |  || |  | |
Db   11751 AAGCACAAGATCCCGCTGGAGTACTGCATCGTCGAGGTGGTCTTCGCCGAGCTCTTTCAT 11692

Qy    1111 ctaccccagccaccattcaagactctttattatacactcgtgattatggatctttgtaag 1170
            | |||  ||| |  |    |   | |  ||      | |  | || || || || ||
Db   11691 ATGCCCACGCCGCGCTATCTGGACATCTGCTACGGCTCCATTTTGATCGAACTCTGCAAA 11632

Qy    1171 gctcttccgggtgcctttcctgctgttgttgctggcgctgttcgtgcactatttgagaaa 1230
             |  ||||   |  | ||   ||| | ||    ||          || |||  |
Db   11631 CTGCAGCCGGCCACGCTGCCCCAGGTTCTGGCCCAGGCCACCGAGATCCTCTTTATGCGC 11572

Qy    1231 atatccgacttagacatggaatccaggacgcgtcttatcctctggttttctcaccactta 1290
           ||   |    | ||| |  ||        ||  |  ||   ||||| ||  | ||  |
Db   11571 ATCGACTCGATGAACACGTCATGTTTCGATCGGTTCGTCAATTGGTTCTCCTATCATCTG 11512

Qy    1291 tccaacttccaattcatctggccgtgggaagagtgggcttttgtgttggatcttcccaag 1350
             ||||||| | ||||  ||| | ||||| || ||||         ||   ||   |
Db   11511 AGCAACTTCAAGTTCACATGGTCATGGGACGAATGGGACAGCTGTCTGCTCCTCGACGGC 11452

Qy    1351 tgggcccctaagcgtgtatttgttcaggagattttgcaaagagaagtacgcttg 1404
             |   ||   |   |||  | |||||| | ||||| |      ||||  | |
Db   11451 GAGCATCCGCGTCCCAAATTCATCCAGGAGGTGTTGCAGAAGTGCCTACGGTAG 11398
```

RESULT 14
AE003431
LOCUS AE003431 311464 bp DNA linear INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386054 section 15 of 35, complete sequence.
ACCESSION AE003431 AE002566
VERSION AE003431.2 GI:10728461
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 311464)
AUTHORS Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,

```
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 311464)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7290491.
FEATURES             Location/Qualifiers
     source          1. .311464
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
     mRNA            complement(<6034. .>6165)
                     /gene="CG12687"
                     /product="CT35554"
                     /db_xref="FLYBASE:FBan0012687"
                     /db_xref="FLYBASE:FBgn0040908"
                     /evidence=not_experimental
     gene            complement(<6034. .>6165)
                     /gene="CG12687"
                     /map="4C1-4C1"
                     /db_xref="FLYBASE:FBan0012687"
                     /db_xref="FLYBASE:FBgn0040908"
                     /evidence=not_experimental
     CDS             complement(6034. .6165)
                     /gene="CG12687"
                     /note="CG12687 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012687"
                     /db_xref="FLYBASE:FBgn0040908"
                     /evidence=not_experimental
                     /protein_id="AAF45945.1"
                     /db_xref="GI:7290492"
                     /translation="MTKCLRFCVGTIPSDSITISPLTTNVATVYHHLSGDLIAGVAG"
     mRNA            join(39063. .40127,61287. .61400,61476. .61638,88782. .88928,
                     88988. .89064,92765. .93018,93151. .94255,94315. .94412)
                     /gene="bi"
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /product="CT12025"
                     /db_xref="FLYBASE:FBan0003578"
                     /db_xref="FLYBASE:FBgn0000179"
     gene            <39063. .>94412
                     /gene="bi"
                     /note="CG3578"
                     /map="4C2-4C3"
                     /db_xref="FLYBASE:FBan0003578"
                     /db_xref="FLYBASE:FBgn0000179"
     CDS             join(39064. .40127,61287. .61400,61476. .61638,88782. .88928,
                     88988. .89064,92765. .93018,93151. .94255,94315. .94357)
                     /gene="bi"
                     /note="bi gene product; Nucleotide sequence of the Celera
                     sequence differs from the published sequence for this
                     transcript"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0003578"
                     /db_xref="FLYBASE:FBgn0000179"
                     /protein_id="AAF45946.1"
                     /db_xref="GI:7290493"
                     /translation="MRYDVQELLFHQSAEDPFARFANGMAYHPFLQLTQRPTDFSVSS
                     LLTAGSNNNNSGNTNSGNNNSNSNNNTNSNTNNTNNLVAVSPTGGGAQLSPQSNHSSS
                     NTTTTSNTNNSSSNNNNNNSTHNNNNNHTNNNNNNNNNTSQKQGHHLSTTEEPPSPAG
                     TPPPTIVGLPPIPPPNNNSSSSSSNNSASAAAHPSHHPTAAHHSPSTGAAAPPAGATG
                     LPPPTPPHHLQQQQQQQQHPAPPPPYFPAAALAALAGSPAGPHPGLYPGGGLRFPPH
                     HPGAHPHAHHLGSAYTTAEDVVLASAVAHQLHPAMRPLRALQPEDDGVVDDPKVTLEG
                     KDLWEKFHKLGTEMVITKSGRQMFPQMKFRVSGLDAKAKYILLLDIVAADDYRYKFHN
                     SRWMVAGKADPEMPKRMYIHPDSPTTGEQWMQKVVSFHKLKLTNNISDKHGFVSTTIL
                     NSMHKYQPRFHLVRANDILKLPYSTFRTYVFKETEFIAVTAYQNEKITQLKIDNNPFA
                     KGFRDTGAGKREKNCYRQALMSNRGSDSDKLNPTHVSSSRAPLHLGHAGRPPHLHPHA
                     ALLDNQQDDDDKLLDVVGPPQSPLLPLSHSLQQMHAHQHSAALAAWFNHLAGAGAGAS
                     EHAAAAAANASAEDALRRRLQADADVERDGSDSSCSESVGGSTGGAFRPTSTGSPKEA
                     VGAAAAAAAAGLNPGGGSYPSPNISVGPPIHPSPHLLPYLYPHGLYPPPHLGLLHNPA
                     AAAAMSPAGLNPGLLFNAQLALAAQHPALFGHAYAAAGHTPVSPLQGLKSHRFSPYSL
                     PGSLGSAFDAVTPGSNANRSGDPPGGGGGGLGGGVVENGPRSLSSSPRPRPASHSPPT
                     RPISMSPTTPPSLMKQPRGGGAGAGVAQSQHSPSELKSMEKMVNGLEVQHNGSAAAAA
                     AALQLAEEAAQHHHHTQAHHQQQQHQSHHQQQHHQQPAQPHPHHQTHLHSHHGATTGG
                     TDQVGGGCRGSGSGSG"
     mRNA            complement(join(47962. .48065,48215. .>48506))
                     /gene="CG12686"
                     /product="CT35553"
                     /db_xref="FLYBASE:FBan0012686"
                     /db_xref="FLYBASE:FBgn0040909"
     gene            complement(<47962. .>48506)
                     /gene="CG12686"
                     /map="4C2-4C2"
                     /db_xref="FLYBASE:FBan0012686"
                     /db_xref="FLYBASE:FBgn0040909"
     CDS             complement(join(47962. .48065,48215. .48233))
                     /gene="CG12686"
                     /note="CG12686 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012686"
                     /db_xref="FLYBASE:FBgn0040909"
                     /protein_id="AAF45947.1"
                     /db_xref="GI:7290494"
                     /translation="MQTDSTLQQEFFGDICTPLTHQPASAHCISSKTSRYPLSM"
     mRNA            join(<126141. .126277,126550. .>126709)
                     /gene="CG12685"
                     /product="CT35551"
                     /db_xref="FLYBASE:FBan0012685"
                     /db_xref="FLYBASE:FBgn0040910"
                     /evidence=not_experimental
     gene            <126141. .>126709
                     /gene="CG12685"
                     /map="4C3-4C4"
                     /db_xref="FLYBASE:FBan0012685"
                     /db_xref="FLYBASE:FBgn0040910"
                     /evidence=not_experimental
     CDS             join(126141. .126277,126550. .126709)
                     /gene="CG12685"
                     /note="CG12685 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012685"
                     /db_xref="FLYBASE:FBgn0040910"
                     /evidence=not_experimental
                     /protein_id="AAF45948.1"
                     /db_xref="GI:7290495"
                     /translation="MDDGAASRAGPETGIWDVRCRMQDAGCWLLAAECWTWKLLLDGL
                     SCAVSAPLVLATQLTRTRQILAPATPTAYRLPPTAPPAQQHHFQWQWAAVGVGK"
     mRNA            join(140773. .141143,141852. .143220,143398. .143593,
                     143662. .143739,143803. .143885,144067. .>144128)
```

```
                     /gene="CG3556"
                     /product="CT11944"
                     /db_xref="FLYBASE:FBan0003556"
                     /db_xref="FLYBASE:FBgn0029708"
     gene            <140773. .>144128
                     /gene="CG3556"
                     /map="4C3-4C4"
                     /db_xref="FLYBASE:FBan0003556"
                     /db_xref="FLYBASE:FBgn0029708"
     CDS             join(141852. .143220,143398. .143593,143662. .143739,
                     143803. .143885,144067. .144128)
                     /gene="CG3556"
                     /note="CG3556 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0003556"
                     /db_xref="FLYBASE:FBgn0029708"
                     /protein_id="AAF45949.1"
                     /db_xref="GI:7290496"
                     /translation="MLSLSSPPWLLLLVLFFFANGSATVVSLAENVTLATLPSTSPIS
                     VSFSTTSSVVAPISTELENQTASSSNLNTNNEASDEQTGNSNSNTSSHSRNINGLPSS
                     NSNIDNANSNSSSVSSLSTTTSAISGVDQQESLILATPTALNGSGTPPEHQTIGQAPP
                     TKGEQEAILRALDWLKEKRASDYGWGNDTHVVILAKELSGGRDPNDSVDGHVQVIQEL
                     EDTLSVKEMEIEILAMLDRHHTLPKPLDLDKLARYVLALGSLCKDPKHFHGHDLVATL
                     QHHEPAQDIEFALTTLSACSSAAHVRKRQIRRLLDIASGVTDQSVDAIAMVILALRCI
                     VTDHRHRHLQHFVRRPARGLATLQDQRGSFGSLRSTALAMQALQDLEYDPAGHWNRTA
                     ASRYILSRQRADGGWSEEPLQDGQEPDIGVGLTADIILALGWKGLGAVRALQCDHVIR
                     ESSDPTENGEPKLAVPFGLSSSAEESDAKNISYTYTLWVGSNVTESFSLSLVSPKNTS
```

Query Match 2.1%; Score 56.8; DB 3; Length 311464;
Best Local Similarity 46.6%; Pred. No. 0.0034;
Matches 221; Conservative 0; Mismatches 247; Indels 6; Gaps 1;

```
Qy    937 aaaatggaggatgtacaaccaattgatcgttttgtcgtggaggagtatttgctggatgtg 996
          ||  ||  || ||  ||  |||| || || ||| |  | ||||||  || | |   || |
Db 262115 AATCTGCCGGGTGCCCACTCAATAGAGCGCTTTTTGATTGAGGAGCATCTACACCATATC 262174

Qy    997 ctcttctatttgaatggatgtcggaaggagtgtgcatcctacatggctaatcttcc---- 1052
          |       |   ||      || ||||| || ||  || |  ||  ||   ||||
Db 262175 ATTGAAACATACCATCACGAGCGAAAGGATTGCGCCGCCCAGCTGCTTAGCTTTCCGTAC 262234

Qy   1053 --tgttacatttcggtacgagtatcttatggcagagacactattttctcagatactgctg 1110
            |   | | |   |||||    || |  |||    | ||  |  || | | |
Db 262235 AAGCACAAGATCCCGCTGGAGTACTGCATCGTCGAGGTGGTCTTCGCCGAGCTCTTTCAT 262294

Qy   1111 ctaccccagccaccattcaagactctttattatacactcgtgattatggatctttgtaag 1170
          | |||  ||| |  |    |    | |  ||      | |  | || || || || ||
Db 262295 ATGCCCACGCCGCGCTATCTGGACATCTGCTACGGCTCCATTTTGATCGAACTCTGCAAA 262354

Qy   1171 gctcttccgggtgcctttcctgctgttgttgctggcgctgttcgtgcactatttgagaaa 1230
             |  ||||   |  | ||    ||| | ||    ||          || |||  |
Db 262355 CTGCAGCCGGCCACGCTGCCCCAGGTTCTGGCCCAGGCCACCGAGATCCTCTTTATGCGC 262414

Qy   1231 atatccgacttagacatggaatccaggacgcgtcttatcctctggtttttctcaccactta 1290
          ||   |    |  ||| |  ||        ||  |  ||   ||||| ||  | ||  |
Db 262415 ATCGACTCGATGAACACGTCATGTTTCGATCGGTTCGTCAATTGGTTCTCCTATCATCTG 262474

Qy   1291 tccaacttccaattcatctggccgtgggaagagtgggcttttgtgttggatcttcccaag 1350
            ||||||| | ||||  ||| | |||||  || ||||          ||   ||   |
Db 262475 AGCAACTTCAAGTTCACATGGTCATGGGACGAATGGGACAGCTGTCTGCTCCTCGACGGC 262534

Qy   1351 tgggcccctaagcgtgtatttgttcaggagattttgcaaagagaagtacgcttg 1404
            |   ||    |   |||  | |||||| | ||||| |      |||| | |
Db 262535 GAGCATCCGCGTCCCAAATTCATCCAGGAGGTGTTGCAGAAGTGCCTACGGTAG 262588
```

```
RESULT 15
AC022016
LOCUS       AC022016              163029 bp    DNA     linear   PRI 14-JUN-2001
DEFINITION  Homo sapiens chromosome 10 clone RP11-165M8, complete sequence.
ACCESSION   AC022016
VERSION     AC022016.6  GI:14423578
KEYWORDS    HTG.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163029)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163029)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 163029)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2001) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Jun 14, 2001 this sequence version replaced gi:13940585.
FEATURES             Location/Qualifiers
     source          1. .163029
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-165M8"
                     /clone_lib="RPCI-11"
BASE COUNT    48008 a  29547 c  30527 g  54947 t
ORIGIN
```

Query Match 2.1%; Score 56.2; DB 9; Length 163029;
Best Local Similarity 56.9%; Pred. No. 0.0048;
Matches 103; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy   1466 ctccaaaagctggtccgaattttatgtattccttggaagaaggtaaagagaaaacagaag 1525
          ||| || || || ||   | |  |||| |    | | |||  || |    |||| |||||
Db  19583 CTCTAACAGGTGCTCTTGAATGAATGTTTCTGATTGGAGATTGTGACATCAAAATAGAAG 19642

Qy   1526 aacagcaattgtcagccgaattgagcaggaaggtcaaggaaaaacaaaccgcacgtgaca 1585
          || | |  |    |  |  | |||| |||| || | ||| | ||||| | ||  ||
Db  19643 AAAAACTTTCAGGACTCACAGAGAGCTGGAACATCCATGAATATCAAACAGAACAGGAAT 19702

Qy   1586 tgatagtgtggattgaagaaacgatatatccagttcatggttttgaagttactcttacaa 1645
          | |  |  |||| |||  |||   | |   |  | ||  ||||||  ||   |||| ||
Db  19703 TAACTGCATGGACTGAATTAACAGAAGACTGAAGTTATCTTTTTGACTTTTTGCTTAAAA 19762

Qy   1646 t 1646
          |
Db  19763 T 19763
```

```
Search completed: August 20, 2002, 14:02:25
Job time: 8759 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 11:35:21 ; Search time 2808.13 Seconds
(without alignments)
13054.143 Million cell updates/sec

Title: US-09-882-986-1
Perfect score: 2716
Sequence: 1 aaagagacgaactgaagaaa.....ctttaaaaaaaaaaaaaaaa 2716

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	590	21.7	591	9	AV555064	AV555064 AV555064
c	2	529.2	19.5	535	9	AV555295	AV555295 AV555295
c	3	442.2	16.3	509	9	AV441539	AV441539 AV441539
	4	417.6	15.4	424	10	BE523479	BE523479 M37F3STM
c	5	413.8	15.2	575	12	BH587474	BH587474 BOHNU40TF
	6	401	14.8	401	9	AV564181	AV564181 AV564181
	7	383.8	14.1	480	10	R90218	R90218 16573 Lambd
	8	347.8	12.8	516	9	AA605490	AA605490 30526 Lam
	9	315.2	11.6	595	9	AW306466	AW306466 se51b03.y
	10	311.8	11.5	682	10	BI935623	BI935623 EST555512
	11	309.4	11.4	663	10	BI933835	BI933835 EST553724
	12	294.2	10.8	643	10	BI935968	BI935968 EST555857
	13	288.2	10.6	637	9	AW267982	AW267982 EST306204
	14	273.2	10.1	466	9	AL365935	AL365935 MtBA03C07
	15	267.8	9.9	632	10	BE590437	BE590437 SB177 Sug
	16	258.6	9.5	585	9	AI894891	AI894891 EST264334
	17	257	9.5	553	10	BI785886	BI785886 sai30c09.
	18	251.2	9.2	593	10	BE402476	BE402476 CSB008D02
	19	246.4	9.1	552	10	BE492845	BE492845 WHE0565_F
	20	243.8	9.0	584	10	BF619837	BF619837 HVSMEc000
	21	240	8.8	240	10	BE530285	BE530285 M77G15STM
	22	237.4	8.7	291	9	AV528587	AV528587 AV528587
	23	232.4	8.6	371	9	AI442604	AI442604 sa43a09.y
	24	226.2	8.3	481	9	AW906575	AW906575 EST342697
	25	221	8.1	458	10	BI935752	BI935752 EST555641
	26	201	7.4	242	9	AV560536	AV560536 AV560536
	27	195.2	7.2	497	10	BG599218	BG599218 EST504118
	28	191.6	7.1	490	9	AW568693	AW568693 si60h04.y
	29	174.6	6.4	895	10	BG368900	BG368900 HVSMEi002
	30	174.2	6.4	466	10	D40981	D40981 RICS3204A R
	31	173.8	6.4	372	10	BE400529	BE400529 AWB004.E0
	32	164.8	6.1	288	10	BI120970	BI120970 F025P93Y
	33	157	5.8	585	10	BJ182040	BJ182040 BJ182040
	34	150	5.5	356	9	AV409428	AV409428 AV409428
	35	145.4	5.4	568	10	BJ188993	BJ188993 BJ188993
c	36	141.8	5.2	794	10	BE824117	BE824117 GM700023A
	37	141.2	5.2	332	10	BE600932	BE600932 PI1_92_H1
	38	137	5.0	210	10	F15394	F15394 ATTS5642 Gr
	39	134.2	4.9	394	10	D41074	D41074 RICS3340A R
	40	127.4	4.7	518	10	BF069198	BF069198 st43e07.y
	41	125.6	4.6	276	10	BM444036	BM444036 EBem09_SQ
	42	123	4.5	424	10	C72488	C72488 C72488 Rice
	43	118	4.3	526	10	BJ177704	BJ177704 BJ177704
c	44	115.2	4.2	714	9	AW348783	AW348783 GM210003A
	45	108.6	4.0	602	10	BJ194498	BJ194498 BJ194498

ALIGNMENTS

RESULT 1
AV555064/c
LOCUS AV555064 591 bp mRNA linear EST 07-SEP-2000
DEFINITION AV555064 Arabidopsis thaliana green siliques Columbia Arabidopsis thaliana cDNA clone SQ003a06F 3', mRNA sequence.
ACCESSION AV555064
VERSION AV555064.1 GI:8726479
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 591)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .591
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ003a06F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 161 a 138 c 117 g 175 t
ORIGIN

Query Match 21.7%; Score 590; DB 9; Length 591;

```
  Best Local Similarity   100.0%;  Pred. No. 1e-102;
  Matches  590;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2111 cagcgaagatgaagcgtttaaaatcaacagtggagaagacaggggaagcggagttatctc 2170
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    591 CAGCGAAGATGAAGCGTTTAAAATCAACAGTGGAGAAGACAGGGGAAGCGGAGTTATCTC 532

Qy   2171 ttcgggagtccctagaggcaaaagaggctcttcttaacagagctctctctgagaccgagg 2230
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 TTCGGGAGTCCCTAGAGGCAAAAGAGGCTCTTCTTAACAGAGCTCTCTCTGAGACCGAGG 472

Qy   2231 ttttactgctcttgctgttccaaagtttcttaggtgtactgaaggaacggctcccagatc 2290
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 TTTTACTGCTCTTGCTGTTCCAAAGTTTCTTAGGTGTACTGAAGGAACGGCTCCCAGATC 412

Qy   2291 caactaaagtgagatcagtgcaggatctaaaatctataggtgctgaagatgacaagccat 2350
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 CAACTAAAGTGAGATCAGTGCAGGATCTAAAATCTATAGGTGCTGAAGATGACAAGCCAT 352

Qy   2351 ctgcgatggacgtggacagcgagaatggaaacccaaagaagagttgcgaagtcggtgaga 2410
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 CTGCGATGGACGTGGACAGCGAGAATGGAAACCCAAAGAAGAGTTGCGAAGTCGGTGAGA 292

Qy   2411 gagaacagtggtgcttatcaacacttggctatctcacggcatttacaaggcaatatgcga 2470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 GAGAACAGTGGTGCTTATCAACACTTGGCTATCTCACGGCATTTACAAGGCAATATGCGA 232

Qy   2471 gcgagatatggcctcacatggagaagttggagtcagaagtgttctcgggtgaagatgtgc 2530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 GCGAGATATGGCCTCACATGGAGAAGTTGGAGTCAGAAGTGTTCTCGGGTGAAGATGTGC 172

Qy   2531 atcctctctttctccaagccatatcttctgcacttcaattcccattacattaatcttcct 2590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    171 ATCCTCTCTTTCTCCAAGCCATATCTTCTGCACTTCAATTCCCATTACATTAATCTTCCT 112

Qy   2591 ctttcaatctcaatcaaacctgtctcttttgtttttttgttatgagattctgattctgaca 2650
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    111 CTTTCAATCTCAATCAAACCTGTCTCTTTTGTTTTTTGTTATGAGATTCTGATTCTGACA 52

Qy   2651 tcaagttattaggaaattgaaaagagtcaaaaaacaagagtttaaacttt 2700
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db     51 TCAAGTTATTAGGAAATTGAAAAGAGTCAAAAAACAAGAGTTTAAACTTT 2
```

```
RESULT   2
AV555295/c
LOCUS       AV555295                535 bp    mRNA    linear   EST 06-SEP-2000
DEFINITION  AV555295 Arabidopsis thaliana green siliques Columbia Arabidopsis
            thaliana cDNA clone SQ011a07F 3', mRNA sequence.
ACCESSION   AV555295
VERSION     AV555295.1  GI:8726710
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SQ011a07F"
                     /clone_lib="Arabidopsis thaliana green siliques Columbia"
                     /tissue_type="green siliques"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      151 a    118 c    111 g    155 t
ORIGIN
```

```
  Query Match             19.5%;  Score 529.2;  DB 9;  Length 535;
  Best Local Similarity   99.4%;  Pred. No. 4.2e-91;
  Matches  531;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy   2170 cttcgggagtccctagaggcaaaagaggctcttcttaacagagctctctctgagaccgag 2229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    535 CTTCGGGAGTCCCTAGAGGCAAAAGAGGCTCTTCTTAACAGAGCTCTCTCTGAGACCGAG 476

Qy   2230 gttttactgctcttgctgttccaaagtttcttaggtgtactgaaggaacggctcccagat 2289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    475 GTTTTACTGCTCTTGCTGTTCCAAAGTTTCTTAGGTGTACTGAAGGAACGGCTCCCAGAT 416

Qy   2290 ccaactaaagtgagatcagtgcaggatctaaaatctataggtgctgaagatgacaagcca 2349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    415 CCAACTAAAGTGAGATCAGTGCAGGATCTAAAATCTATAGGTGCTGAAGATGACAAGCCA 356

Qy   2350 tctgcgatggacgtggacagcgagaatggaaacccaaagaagagttgcgaagtcggtgag 2409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    355 TCTGCGATGGACGTGGACAGCGAGAATGGAAACCCAAAGAAGAGTTGCGAAGTCGGTGAG 296

Qy   2410 agagaacagtggtgcttatcaacacttggctatctcacggcatttacaaggcaatatgcg 2469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    295 AGAGAACAGTGGTGCTTATCAACACTTGGCTATCTCACGGCATTTACAAGGCAATATGCG 236

Qy   2470 agcgagatatggcctcacatggagaagttggagtcagaagtgttctcgggtgaagatgtg 2529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    235 AGCGAGATATGGCCTCACATGGAGAAGTTGGAGTCAGAAGTGTTCTCGGGTGAAGATGTG 176

Qy   2530 catcctctctttctccaagccatatcttctgcacttcaattcccattacattaatcttcc 2589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    175 CATCCTCTCTTTCTCCAAGCCATATCTTCTGCACTTCAATTCCCATTACATTAATCTTCC 116

Qy   2590 tctttcaatctcaatcaaacctgtctcttttgtttttttgttatgagattctgattctgac 2649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    115 TCTTTCAATCTCAATCAAACCTGTCTCTTTTGTTTTTTGTTATGAGATTCTGATTCTGAC 56

Qy   2650 atcaagttattaggaaattgaaaagagtcaaaaaacaagagtttaaacttttaaa 2703
          |||||||| |||||||||||||||||||||||||| |||||||||||| ||
Db     55 ATCAAGTTCTTAGGAAATTGAAAAGAGTCAAAAAACAAAAGTTTAAACTTTTAA 2
```

```
RESULT   3
AV441539/c
LOCUS       AV441539                509 bp    mRNA    linear   EST 14-NOV-2000
DEFINITION  AV441539 Arabidopsis thaliana above-ground organ two to six-week
            old Arabidopsis thaliana cDNA clone APZ63f01_f 3', mRNA sequence.
ACCESSION   AV441539
VERSION     AV441539.1  GI:7611929
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
```

```
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES              Location/Qualifiers
     source           1. .509
                      /organism="Arabidopsis thaliana"
                      /strain="Columbia"
                      /db_xref="taxon:3702"
                      /clone="APZ63f01_f"
                      /clone_lib="Arabidopsis thaliana above-ground organ two to
                      six-week old"
                      /tissue_type="aboveground organs"
                      /dev_stage="two to six-week old"
                      /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                      XhoI"
BASE COUNT      141 a    109 c    104 g    155 t
ORIGIN


  Query Match            16.3%;  Score 442.2;  DB 9;  Length 509;
  Best Local Similarity   93.3%;  Pred. No. 1.7e-74;
  Matches  473;  Conservative    0;  Mismatches   33;  Indels    1;  Gaps    1;

Qy   2185 gaggcaaaagaggctcttcttaacagagctctctctgagaccgaggttttactgctcttg 2244
           |||||||||||||||||||||||||||||||||||||||||| ||  | ||||||||| |||||||
Db    509 GAGGCAAAAGAGGCTCTTCTTAACAGAGCTCTCTCTAAGCACAAGGTTTTACAGCTCTTG 450

Qy   2245 ctgttccaaagtttcttaggtgtactgaaggaacggctcccagatccaactaaagtgaga 2304
           |||| ||||||||||||||||||||| ||||| ||||||||||||||||||||||||||
Db    449 CTGTCCCAAAGTTTCTTAGGTGTACTAAAGGATCGGCTCCCAGATCCAACTAAAGTGAGA 390

Qy   2305 tcagtgcaggatctaaaatctataggtgctgaagatgacaagccatctgcgatggacgtg 2364
           |||||||||||||||||||||||||||||||||||||||  ||||||||| ||||||||||
Db    389 TCAGTGCAGGATCTAAAATCTATAGGTGCTGAAGATGCAAAGCCATCTTCGATGGACGTG 330

Qy   2365 gacagcgagaa-tggaaacccaaagaagagttgcgaagtcggtgagagagaacagtggtg 2423
           |||||||||||| |||||||||||  ||||||   |||||||||||||| ||||||||||
Db    329 GACAGCGAGAATTGGAAACCCAAAATAGAGTTCAAAAGTCGGTGAGAGAAAACAGTGGTG 270

Qy   2424 cttatcaacacttggctatctcacggcatttacaaggcaatatgcgagcgagatatggcc 2483
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    269 CTTATCAACACTTGGCTATCTCACGGCATTTACAAGGCAATATGCGAGCAAGATATGGCC 210

Qy   2484 tcacatggagaagttggagtcagaagtgttctcgggtgaagatgtgcatcctctctttct 2543
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    209 TCACATGGAGAAGTTGGAGTCAAAAGTGTTCTCGGGTGAAGATGTGCATCCTCTCTTTCT 150

Qy   2544 ccaagccatatcttctgcacttcaattcccattacattaatcttcctctttcaatctcaa 2603
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    149 CCAAGCCATATCTTCTGCACTTCAATTCCCATAACATTAATCTTCCTCTTTCAATCTCAA 90

Qy   2604 tcaaacctgtctcttttgttttttgttatgagattctgattctgacatcaagttattagg 2663
           |||||||||||||||||||||||||||||||||| || ||||| ||||||| |||||||
Db     89 TCAAACCTGTCTCTTTTGTTTTTTGTTATGAGATTGTGGTTCTGGCATCAAGATATTAGG 30

Qy   2664 aaattgaaaagagtcaaaaaacaagag 2690
           ||||||||||    ||| | |  ||
Db     29 CAATTGAAAAGATGTCAAAGAAACAAG 3


RESULT   4
BE523479
LOCUS       BE523479                424 bp    mRNA    linear   EST 19-MAR-2001
DEFINITION  M37F3STM Arabidopsis developing seed Arabidopsis thaliana cDNA
            clone M37F3 5', mRNA sequence.
ACCESSION   BE523479
VERSION     BE523479.1  GI:9781457
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
            Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
  TITLE     A new set of Arabidopsis expressed sequence tags from developing
            seeds. The metabolic pathway from carbohydrates to seed oil
  JOURNAL   Plant Physiol. 124 (4), 1582-1594 (2000)
  MEDLINE   20567808
COMMENT     Contact: Benning, C
            Dept. of Biochemistry & Molecular Biology
            Michigan State University
            224 Biochemistry, Michigan State University, East Lansing, MI 48824
            , USA
            Tel: 517 355 1609
            Fax: 517 353 9334
            Email: benning@msu.edu
            Michigan State University DNA Sequencing Facility Arabidopsis
            Biological Resource Center, The Ohio State University, 309 Botany &
            Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
            6142920603 TEL: 6142929371.
FEATURES              Location/Qualifiers
     source           1. .424
                      /organism="Arabidopsis thaliana"
                      /strain="Columbia"
                      /db_xref="taxon:3702"
                      /clone="M37F3"
                      /clone_lib="Arabidopsis developing seed"
                      /tissue_type="seed"
                      /dev_stage="5-13 days after flowering"
                      /lab_host="E.coli"
                      /note="Organ: Developing seed; Vector: pBluescript SK-;
                      Site_1: EcoRI; Site_2: XhoII"
BASE COUNT      105 a     93 c     92 g    134 t
ORIGIN


  Query Match            15.4%;  Score 417.6;  DB 10;  Length 424;
  Best Local Similarity   99.1%;  Pred. No. 8.8e-70;
  Matches  420;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy     24 ctctcggaagaagatgagcaattggaaaactcttctccttcgcatcggcgaaaagggacc 83
          | | ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGCACGAGAGAAGATGAGCAATTGGAAAACTCTTCTCCTTCGCATCGGCGAAAAGGGACC 60

Qy     84 tgagtacggcacttcctccgactacaaagaccacatcgagacttgtttcggtgtcattcg 143
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGAGTACGGCACTTCCTCCGACTACAAAGACCACATCGAGACTTGTTTCGGTGTCATTCG 120

Qy    144 tagagaaatcgagcgttctggagatcaagttttgccttttctactacaatgtgctgaaca 203
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TAGAGAAATCGAGCGTTCTGGAGATCAAGTTTTGCCTTTTCTACTACAATGTGCTGAACA 180

Qy    204 attgcctcataagattcctttgtatgggactttgattggtttgttgaacttggagaatga 263
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATTGCCTCATAAGATTCCTTTGTATGGGACTTTGATTGGTTTGTTGAACTTGGAGAATGA 240

Qy    264 agattttgtccagaagctagtagaaagtgtccacgctaatttccaggtcgctttagattc 323
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGATTTTGTCCAGAAGCTAGTAGAAAGTGTCCACGCTAATTTCCAGGTCGCTTTAGATTC 300

Qy    324 tggcaactgcaacagtatccgtatattgcttcgctttatgacttccctgttgtgcagtaa 383
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGGCAACTGCAACAGTATCCGTATATTGCTTCGCTTTATGACTTCCCTGTTGTGCAGTAA 360

Qy    384 ggttattcaacctgcttctttgattgtcgtcttcgaaacattgctatcatctgctgccac 443
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGTTATTCAACCTGCTTCTTTGATTGTCGTCTTCGAAACATTGCTATCATCTGCTGCCAC 420

Qy    444 tact 447
          ||||
Db    421 TACT 424


RESULT   5
```

```
BH587474/c
LOCUS       BH587474                575 bp    DNA     linear   GSS 15-DEC-2001
DEFINITION  BOHNU40TF BOHN Brassica oleracea genomic clone BOHNU40, DNA
            sequence.
ACCESSION   BH587474
VERSION     BH587474.1  GI:17839932
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOHNU40TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .575
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHNU40"
                     /clone_lib="BOHN"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      157 a    148 c    102 g    168 t
ORIGIN
```

```
  Query Match             15.2%;  Score 413.8;  DB 12;  Length 575;
  Best Local Similarity   87.1%;  Pred. No. 4.3e-69;
  Matches  454;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy   1421 ttaagcagagcattgagaatgcgactgccctagaagaattacttcctccaaaagctggtc 1480
          | ||  ||||||| |||||||| | ||||||||||||||| |||||| |||| |||| ||
Db    521 TCAAATAGAGCATCGAGAATGCTAGTGCCCTAGAAGAATTGCTTCCTGCAAAGGCTGCTC 462

Qy   1481 cgaattttatgtattccttggaagaaggtaaagagaaaacagaagaacagcaattgtcag 1540
          ||  ||  | ||| |||||||||| |||||||||||||||||||| || || |||||||
Db    461 CGGCTTACAGGTACTCCTTGGAAGAGGGTAAAGAGAAAACAGAAGAGCATCAGTTGTCAG 402

Qy   1541 ccgaattgagcaggaaggtcaaggaaaaacaaaccgcacgtgacatgatagtgtggattg 1600
          | ||| ||| |||||||||||||||||||||| | ||||||||||||||    ||||||||
Db    401 CTGAACTGAACAGGAAGGTCAAGGAAAAACAATCTGCACGTGACATGATGTCGTGGATTG 342

Qy   1601 aagaaacgatatatccagttcatggttttgaagttactcttacaatagttgtacagacct 1660
          |||||||||||||||||||||||||||||||| ||||||||| |||| || |||||||
Db    341 AAGAAACGATATATCCAGTTCATGGTTTTGAAGTCACTCTTACAGTAGTCGTGCAGACCT 282

Qy   1661 tacttgacatcggatcaaaaagtttcactcatttggtcactgtcctggagcgatatggcc 1720
          |||| || |||||||||||||||||||||||| ||||||| ||||| ||||| || ||||
Db    281 TACTCGAGATCGGATCAAAAAGTTTCACTCATATGGTCACAGTCCTAGAGCGGTACGGCC 222

Qy   1721 aagtattttcaaagctttgtcctgataacgataagcaggtgatgctattatctcaagtga 1780
          |||| |||   || || |||||||||| ||||| ||||||||||||||| |||||||
Db    221 AAGTGTTTGGGAAACTCTGTCCTGATAATGATAAACAGGTGATGCTATTATCGCAAGTGA 162

Qy   1781 gtacatactggaaaaacaatgtacaaatgacggcggtggcaattgataggatgatgggtt 1840
          || |||||||||||||||| |||||||||| |||||||| |||||||||||||| |
Db    161 GTGCATACTGGAAAAACAATGCACAAATGACGGCCGTGGCAATGGATAGGATGATGGGGT 102

Qy   1841 atagactagtatctaatcaggcaattgttagatgggtgttctctccagaaaatgttgatc 1900
          ||||||| ||||| |||||||||||||| |||||||| |||||||| || ||||||||||
```

```
Db    101 ATAGACTCGTATCGAATCAGGCAATTGTGAGATGGGTTTTCTCTCCGGAGAATGTTGATC 42

Qy   1901 agtttcatgtgtctgatcagccatgggagatacttggcaat 1941
          |||||||| |||||||||| |||||||| |   | ||| |
Db     41 AGTTTCATATGTCTGATCAGACATGGGAGGTCAGTTGCATT 1
```

```
RESULT   6
AV564181
LOCUS       AV564181                401 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  AV564181 Arabidopsis thaliana green siliques Columbia Arabidopsis
            thaliana cDNA clone SQ200c06F 3', mRNA sequence.
ACCESSION   AV564181
VERSION     AV564181.1  GI:8735607
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .401
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SQ200c06F"
                     /clone_lib="Arabidopsis thaliana green siliques Columbia"
                     /tissue_type="green siliques"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      135 a     71 c     88 g    107 t
ORIGIN
```

```
  Query Match             14.8%;  Score 401;  DB 9;  Length 401;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-66;
  Matches  401;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1397 tacgcttgtcttactgggataaaattaagcagagcattgagaatgcgactgccctagaag 1456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TACGCTTGTCTTACTGGGATAAAATTAAGCAGAGCATTGAGAATGCGACTGCCCTAGAAG 60

Qy   1457 aattacttcctccaaaagctggtccgaattttatgtattccttggaagaaggtaaagaga 1516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AATTACTTCCTCCAAAAGCTGGTCCGAATTTTATGTATTCCTTGGAAGAAGGTAAAGAGA 120

Qy   1517 aaacagaagaacagcaattgtcagccgaattgagcaggaaggtcaaggaaaaacaaaccg 1576
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAACAGAAGAACAGCAATTGTCAGCCGAATTGAGCAGGAAGGTCAAGGAAAAACAAACCG 180

Qy   1577 cacgtgacatgatagtgtggattgaagaaacgatatatccagttcatggttttgaagtta 1636
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CACGTGACATGATAGTGTGGATTGAAGAAACGATATATCCAGTTCATGGTTTTGAAGTTA 240

Qy   1637 ctcttacaatagttgtacagaccttacttgacatcggatcaaaaagtttcactcatttgg 1696
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCTTACAATAGTTGTACAGACCTTACTTGACATCGGATCAAAAAGTTTCACTCATTTGG 300

Qy   1697 tcactgtcctggagcgatatggccaagtattttcaaagctttgtcctgataacgataagc 1756
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCACTGTCCTGGAGCGATATGGCCAAGTATTTTCAAAGCTTTGTCCTGATAACGATAAGC 360
```

```
Qy   1757 aggtgatgctattatctcaagtgagtacatactggaaaaac 1797
          ||||||||||||||||||||||||||||||||||||||||||
Db    361 AGGTGATGCTATTATCTCAAGTGAGTACATACTGGAAAAAC 401


RESULT   7
R90218
LOCUS       R90218                   480 bp    mRNA    linear   EST 30-DEC-1997
DEFINITION  16573 Lambda-PRL2 Arabidopsis thaliana cDNA clone 191I8T7, mRNA
            sequence.
ACCESSION   R90218
VERSION     R90218.1  GI:957758
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
            ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
            ,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7 dye primer.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Arabidopsis thaliana"
                     /strain="var columbia"
                     /db_xref="taxon:3702"
                     /clone="191I8T7"
                     /clone_lib="Lambda-PRL2"
                     /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
                     Lambda PRL2 is a cDNA library derived from equal
                     quantities of 4 pools of mRNA.  The mRNA sources were 1) 7
                     day germinated etiolated seedlings; 2) tissue culture
                     grown roots; 3) staged plants half with 24 hour light
                     cycle, half on 16 hr light, 8 hour dark- rosettes;  4)
                     same plants as 3 but aerial tissue (stems, flowers and
                     siliques.  The vector is BRL's lambda Zip-Lox.  The cDNA
                     inserts were directionally cloned with Sal-Not arms using
                     oligo dT primed cDNA.  "
BASE COUNT      124 a     80 c    125 g    132 t     19 others
ORIGIN


  Query Match           14.1%;  Score 383.8;  DB 10;  Length 480;
  Best Local Similarity   91.5%;  Pred. No. 2.4e-63;
  Matches  420;  Conservative    0;  Mismatches   36;  Indels    3;  Gaps    2;

Qy   1763 tgctattatctcaagtgagtacatactggaaaaacaatgtacaaatgacggcggtggcaa 1822
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGCTATTATCTCAAGTGAGTACATACTGGAAAAACAATGTACAAATGACGGCGGTGGCAA 60

Qy   1823 ttgataggatgatgggttatagactagtatctaatcaggcaattgttagatgggtgttct 1882
          |||||||||||||||||||||||||||||||||||||||| |||||||||| ||
Db     61 TTGATAGGATGATGGGTTATAGACTAGTATCTAATCAGGCAATTGTNAGATGGGTGTNCT 120

Qy   1883 ctccagaaaatgttgatcagtttcatgtgtctgatcagccatgggagatacttggcaatg 1942
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTCCAGAAAATGTTGATCAGTTTCATGTGTCTGATCAGCCATGGGAGATACTTGGCAATG 180

Qy   1943 ctcttaacaagacttataaccgtatctctgatttgaggaaagatatatcaaacattacga 2002
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     181 CTCTTAACAAGACTTATAACCGTATCTCTNATTTGAGGAAAGATATATCAAACATTACGA 240

Qy   2003 aaaatgttttggttgctgagaaagcttcagccaatgcacgagtagagttggaggctgctg 2062
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AAAATGTTTTGGTTGCTGAGAAAGCTTCAGCCAATGCACGAGTAGAGTTGGAGGCTGCTG 300

Qy   2063 agagcaaactttccctagtggaaggtgaacccgttcttggtgagaatccagcgaagatga 2122
          |||||||||||||||||||| |||   ||||||||||||||||||||||| ||||||
Db     301 AGAGCAAACTTTCCCTAGTGGNAGGGTGACCCGTTCTTGGTGAGAATCCAGCGNAGATGA 360

Qy   2123 agcgtttaaaatcaac--agtggagaagacaggggaagc-ggagttatctcttcgggagt 2179
          ||||||| |||   ||   |||| || | |||||| ||| || ||| ||| |||||| |
Db     361 AGCGTTTTAAAATCACCAGGTGGNGANGCCAGGGGNAGCGGGGGTTTTCTTTTCGGGNGC 420

Qy   2180 ccctagaggcaaaagaggctcttcttaacagagctctct 2218
          |||| | |||||||  || |||| |||   |   | | |
Db     421 CCCTNGGGGCAAAANGGGNTCTTTTTACAGGNCTTTTNT 459


RESULT   8
AA605490
LOCUS       AA605490                 516 bp    mRNA    linear   EST 29-SEP-1997
DEFINITION  30526 Lambda-PRL2 Arabidopsis thaliana cDNA clone 191I8XP 3', mRNA
            sequence.
ACCESSION   AA605490
VERSION     AA605490.1  GI:2446018
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
            ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
            ,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            The sequence entry for this EST has been reverse complimented and
            is being submitted in the sense orientation.
            Seq primer: M13 -21 dye primer.
FEATURES             Location/Qualifiers
     source          1. .516
                     /organism="Arabidopsis thaliana"
                     /strain="var columbia"
                     /db_xref="taxon:3702"
                     /clone="191I8XP"
                     /clone_lib="Lambda-PRL2"
                     /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
                     Lambda PRL2 is a cDNA library derived from equal
                     quantities of 4 pools of mRNA.  The mRNA sources were 1) 7
                     day germinated etiolated seedlings; 2) tissue culture
                     grown roots; 3) staged plants half with 24 hour light
                     cycle, half on 16 hr light, 8 hour dark- rosettes;  4)
                     same plants as 3 but aerial tissue (stems, flowers and
                     siliques.  The vector is BRL's lambda Zip-Lox.  The cDNA
                     inserts were directionally cloned with Sal-Not arms using
                     oligo dT primed cDNA.  "
BASE COUNT      126 a    114 c    118 g    138 t     20 others
ORIGIN
```

```
  Query Match            12.8%;  Score 347.8;  DB 9;  Length 516;
  Best Local Similarity  89.0%;  Pred. No. 1.7e-56;
  Matches  414;  Conservative    0;  Mismatches   44;  Indels    7;  Gaps    4;

Qy   2239 ctcttgctgttccaaagtttcttaggtgtactgaaggaacggctcccagatccaactaaa 2298
          | |||||  | |||||||||| ||  |  | || | ||| || | |||  |||| ||||||
Db     55 CCCTTGCGTTCCCAAAGTTTTTT--GGNTTCTTAGGGACCGNCNCCCNAATCCCACTAAA 112

Qy   2299 gtgagatcagtgcaggatctaaaatctataggtgctgaagatgac-aagccatctgcgat 2357
          | |    | |  | ||   | |||||| |||| |||||||||  | ||  |  |||||||
Db    113 GNGGACCCGGGCCGGG--TTTAAATCTNTAGGNGCTGAAGATTNCCAANGCCNCTGCGAT 170

Qy   2358 ggacgtggacagcgagaatggaaaccc--aaagaagagttgcgaagtcggtgagagagaa 2415
          ||||  ||||||||  | |  | |||||  || ||||||||||||||||||||||||||
Db    171 GGACNTGGACAGCCGGGAATGGAACCCCAAAGGAAGAGTTGCGAAGTCGGTGAGAGAGAA 230

Qy   2416 cagtggtgcttatcaacacttggctatctcacggcatttacaaggcaatatgcgagcgag 2475
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 CAGTGGTGCTTATCAACACTTGGCTATCTCACGGCATTTACAAGGCAATATGCGAGCGAG 290

Qy   2476 atatggcctcacatggagaagttggagtcagaagtgttctcgggtgaagatgtgcatcct 2535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 ATATGGCCTCACATGGAGAAGTTGGAGTCAGAAGTGTTCTCGGGTGAAGATGTGCATCCT 350

Qy   2536 ctctttctccaagccatatcttctgcacttcaattcccattacattaatcttcctctttc 2595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 CTCTTTCTCCAAGCCATATCTTCTGCACTTCAATTCCCATTACATTAATCTTCCTCTTTC 410

Qy   2596 aatctcaatcaaacctgtctcttttgtttttttgttatgagattctgattctgacatcaag 2655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 AATCTCAATCAAACCTGTCTCTTTTGTTTTTTGTTATGAGATTCTGATTCTGACATCAAG 470

Qy   2656 ttattaggaaattgaaaagagtcaaaaaacaagagtttaaacttt 2700
          |||||||||||||||||||||||||||||| ||||||||||
Db    471 TTATTAGGAAATTGAAAAGAGTCAAAAAACAAGNGTTTAAACTTT 515


RESULT   9
AW306466
LOCUS       AW306466                595 bp    mRNA    linear   EST 02-DEC-2001
DEFINITION  se51b03.y1 Gm-c1017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1017-2406 5' similar to SW:CB80_HUMAN Q09161 80 KD NUCLEAR CAP
            BINDING PROTEIN ;, mRNA sequence.
ACCESSION   AW306466
VERSION     AW306466.1  GI:6718819
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 595)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 768   Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 417.
FEATURES             Location/Qualifiers
     source          1. .595
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1017-2406"
                     /clone_lib="Gm-c1017"
                     /tissue_type="vegetable buds of field grown plants"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from vegatable buds of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
BASE COUNT      182 a     99 c    147 g    167 t
ORIGIN


  Query Match            11.6%;  Score 315.2;  DB 9;  Length 595;
  Best Local Similarity  72.0%;  Pred. No. 2.8e-50;
  Matches  426;  Conservative    0;  Mismatches  163;  Indels    3;  Gaps    1;

Qy   1376 aggagattttgcaaagagaagtacgcttgtcttactgggataaaattaagcagagcattg 1435
          ||||| ||||| | || |||||| |||||||| |||||||| ||| ||||||||||| ||||
Db      6 AGGAGGTTTTGGAGAGGGAAGTTCGCTTGTCATACTGGGACAAAGTTAAGCAGAGTATTG 65

Qy   1436 agaatgcgactgccctagaagaattacttcctccaaaagctggtccgaatttttatgtatt 1495
          | |||||  |||   |||||||||| |||||||||| | ||| ||||| |||| | | |
Db     66 AAAATGCACCTGGTTTAGAAGAATTGCTTCCTCCAAAGGGTGGGCCGAACTTTAGTTTTG 125

Qy   1496 ccttggaagaaggtaaagagaaaacagaagaacagcaattgtcagccgaattgagcagga 1555
               ||||| | |||||   |||   | |||||   | |||| |   |  | | ||  |
Db    126 GTGCAGAAGATGATAAAG---AAAGCAATGAACATGTACTGTCTGGACAGCTCAACAACA 182

Qy   1556 aggtcaaggaaaaacaaaccgcacgtgacatgatagtgtggattgaagaaacgatatatc 1615
          ||| || | |||   | | |  ||||| || ||    |||||||| ||||  |   ||
Db    183 TGGTTAAAGGAAAGGCACCTGTCCGTGAAATAATCTCATGGATTGATGAAAGTGTGCTTC 242

Qy   1616 cagttcatggttttgaagttactcttacaatagttgtacagaccttacttgacatcggat 1675
          ||  | ||||| | |||||||| || | | | |||||||| ||  |  ||  | || ||||
Db    243 CAAATAATGGTCTAGAAGTTACCCTCAGAGTGGTTGTACAAACTCTTCTCAATATTGGAT 302

Qy   1676 caaaaagtttcactcatttggtcactgtcctggagcgatatggccaagtattttcaaagc 1735
          | || |||||||||||||| | |||||| ||||| ||||||| ||||| ||| ||||
Db    303 CCAAGAGTTTCACTCATTTGATGACTGTCTTGGAGAGATATGGGCAAGTCTTTGCAAAAT 362

Qy   1736 tttgtcctgataacgataagcaggtgatgctattatctcaagtgagtacatactggaaaa 1795
          | ||||||||  | |||||||| || |||||  || || |||||||| | | ||||||
Db    363 TATGTCCTGACCAGGATAAGCAAGTCATGCTGATAGCTGAAGTGAGTTCTTTCTGGAAGA 422

Qy   1796 acaatgtacaaatgacggcggtggcaattgataggatgatgggttatagactagtatcta 1855
            |||   || |||||  ||  | || ||||| ||||||||||||||||   || || || |
Db    423 GTAATACCCAGATGACAGCAATAGCTATTGACAGGATGATGGGTTATCAGCTTGTTTCAA 482

Qy   1856 atcaggcaattgttagatgggtgttctctccagaaaatgttgatcagtttcatgtgtctg 1915
          ||| ||| ||||| |||||||| |||||  |||| ||| |||| || |||||| |||| |
Db    483 ATCTGGCTATTGTGAGATGGGTCTTCTCAGCAGAGAATATTGAGCAATTTCATATGTCAG 542

Qy   1916 atcagccatgggagatacttggcaatgctcttaacaagacttataaccgtat 1967
          |||   | | |||||| ||| | |||||  | | ||| ||  | || |||||
Db    543 ATCGTGCCTTGGAGATTCTTAGAAATGCAGTAAGCAACACACACAATCGTAT 594
```

```
RESULT  10
BI935623
LOCUS       BI935623                682 bp    mRNA    linear   EST 18-OCT-2001
DEFINITION  EST555512 tomato flower, anthesis Lycopersicon esculentum cDNA
            clone cTOD23B17 5' end, mRNA sequence.
ACCESSION   BI935623
VERSION     BI935623.1  GI:16250095
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 682)
  AUTHORS   van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
            Utterback,T., Van Aken,S., Ronning,C.M., Nierman,W., Fraser,C.M.,
            Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
  TITLE     Generation of ESTs from tomato flower tissue, anthesis (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            This clone is available through the Clemson University Genomics
            Institute
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .682
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cTOD23B17"
                     /clone_lib="tomato flower, anthesis"
                     /tissue_type="flower"
                     /dev_stage="anthesis"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; supplier: Cornell University; sequencing: The
                     Institute for Genomic Research; Flower buds and flowers
                     were taken from greenhouse plants (4-8 wks old, TA496).
                     They were immediately frozen in liquid nitrogen and then
                     size-separated while remaining frozen."
BASE COUNT      196 a    121 c    171 g    194 t
ORIGIN


  Query Match             11.5%;  Score 311.8;  DB 10;  Length 682;
  Best Local Similarity   68.9%;  Pred. No. 1.2e-49;
  Matches  457;  Conservative    0;  Mismatches  202;  Indels    4;  Gaps    2;

Qy   1445 ctgccctagaagaattacttcctccaaaagctggtccgaattttatgtattccttggaag 1504
           ||||  | ||  ||  |||||||  |||| || ||| ||  | ||||  |||     |||||
Db      1 CTGCTTTGGAGGAGCTACTTCCACCAAGAGGTGGACCACAATTTAAATATAGTGCGGAAG 60

Qy   1505 aaggtaaagagaaaacagaagaacagcaattgtcagccgaattgagcaggaaggtcaagg 1564
           | ||||  ||   ||| ||    |   | | ||   ||| | |     | ||| || |
Db     61 ATGGTACTGATCCAACTGAG---CGTGCACTTTCTTTAGAACTCAAGGACATGGTGAAAG 117

Qy   1565 aaaaacaaaccgcacgtgacatgatagtgtggattgaagaaacgatatatccagttcatg 1624
             |   |||| || || || |||||    ||| | |||||||   | | |||||  |||| 
Db    118 GTAGGAAAACTGCGCGGGAAATGATCTCCTGGGTAGAAGAAAATGTTTTTCCAGCCCATG 177

Qy   1625 gttttgaagttactcttacaatagttgtacagaccttacttgacatcggatcaaaaagtt 1684
           | |||||  |||| |||  | | ||||| || ||  | || ||||||||||||||||| |
Db    178 GGTTTGATATTACCCTTGGAGTTGTTGTTCAAACACTTCTCGACATCGGATCAAAAAGCT 237

Qy   1685 tcactcatttggtcactgtcctggagcgatatggccaagtattttcaaagctttgtcctg 1744
           | |||||||| | |||||| ||||| ||||||||||||  || ||||| | ||| |||
Db    238 TTACTCATTTGATTACTGTCTTGGAGAGATATGGCCAAGTCATTGCAAAGATGTGTACTG 297

Qy   1745 ataacgataagcaggtgatgctattatctcaagtgagtacatactggaaaaacaatgtac 1804
           || | ||| | || || | |||  |  |  |||| ||| | |||||| |||||| ||  |

Db    298 ATGATGATCAACAAGTTAAGCTGATTACGGAAGTAAGTTCTTACTGGCAAAACAGTGCGC 357

Qy   1805 aaatgacggcggtggcaattgataggatgatgggttatagactagtatctaatcaggcaa 1864
           ||||||| ||  | || || ||  |||||||| ||||| | ||  |||||||  ||| |
Db    358 AAATGACAGCTATTGCCATCGACCGGATGATGAGTTATCGGCTTATATCTAATTTGGCCA 417

Qy   1865 ttgttagatgggtgttctctccagaaaatgttgatcagtttcatgtgtctgatcagccat 1924
           | || |||||||| |||||||||   ||| |||||| |||||||| |||||     | |
Db    418 TAGTGAGATGGGTTTTCTCTCCACTGAATCTTGATCGGTTTCATGTTTCTGACAGTTCGT 477

Qy   1925 gggagatacttggcaatgctcttaacaagacttataaccgtatctctgatttgaggaaag 1984
           ||||||| ||  | |||||| | |  ||||| ||||| || || |||||| |||| ||||
Db    478 GGGAGATCCTCAGGAATGCTGTCAGTAAGACGTATAATCGGATTTCTGATCTGAGAAAAG 537

Qy   1985 atatatcaaacattacgaaaaatgttttggttgctgagaaagcttcagccaatgcacgag 2044
           | || |||  | ||  || || |||| |  | ||||| |||||| || | |  ||  |||
Db    538 AGATTTCATCCCTTGAGAGAAGTGTTGTCTTGGCTGA-AAAGCTGCATCTAGAGCCAGAG 596

Qy   2045 tagagttggaggctgctgagagcaaactttccctagtggaaggtgaacccgttcttggtg 2104
            |||||| ||  |||| |||   || ||     |  | || || ||||| ||||||||||
Db    597 AAGAGTTAGAATCTGCCGAGTCTAAGCTGAGTGTCATAGATGGGGAACCTGTTCTTGGTG 656

Qy   2105 aga 2107
           | | |
Db    657 AAA 659


RESULT  11
BI933835
LOCUS       BI933835                663 bp    mRNA    linear   EST 18-OCT-2001
DEFINITION  EST553724 tomato flower, anthesis Lycopersicon esculentum cDNA
            clone cTOD18E1 5' end, mRNA sequence.
ACCESSION   BI933835
VERSION     BI933835.1  GI:16248307
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
            Utterback,T., Van Aken,S., Ronning,C.M., Nierman,W., Fraser,C.M.,
            Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
  TITLE     Generation of ESTs from tomato flower tissue, anthesis (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            This clone is available through the Clemson University Genomics
            Institute
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cTOD18E1"
                     /clone_lib="tomato flower, anthesis"
                     /tissue_type="flower"
                     /dev_stage="anthesis"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; supplier: Cornell University; sequencing: The
                     Institute for Genomic Research; Flower buds and flowers
                     were taken from greenhouse plants (4-8 wks old, TA496).
                     They were immediately frozen in liquid nitrogen and then
                     size-separated while remaining frozen."
BASE COUNT      186 a    120 c    167 g    190 t
ORIGIN
```

```
  Query Match           11.4%;  Score 309.4;  DB 10;  Length 663;
  Best Local Similarity  68.5%;  Pred. No. 3.5e-49;
  Matches  457;  Conservative    0;  Mismatches  206;  Indels    4;  Gaps    2;

Qy   1445 ctgccctagaagaattacttcctccaaaagctggtccgaattttatgtattccttggaag 1504
          ||||  | || ||  ||||||| |||| || ||| ||  | ||||  |||     |||||
Db      1 CTGCTTTGGAGGAGCTACTTCCACCAAGAGGTGGACCACAATTTAAATATAGTGCGGAAG 60

Qy   1505 aaggtaaagagaaaacagaagaacagcaattgtcagccgaattgagcaggaaggtcaagg 1564
          | ||||  ||   ||| ||    |    | | ||    ||| | |     | ||| || |
Db     61 ATGGTACTGATCCAACTGAG---CGTGCACTTTCTTTAGAACTCAAGGACATGGTGAAAG 117

Qy   1565 aaaaacaaaccgcacgtgacatgatagtgtggattgaagaaacgatatatccagttcatg 1624
            |   |||| || || || |||||    ||| | |||||||   | | |||||  ||||
Db    118 GTAGGAAAACTGCGCGGGAAATGATCTCCTGGGTAGAAGAAAATGTTTTTCCAGCCCATG 177

Qy   1625 gttttgaagttactcttacaatagttgtacagaccttacttgacatcggatcaaaaagtt 1684
          | |||||  |||| |||  | | ||||| || ||  | || ||||||||||||||||| |
Db    178 GGTTTGATATTACCCTTGGAGTTGTTGTTCAAACACTTCTCGACATCGGATCAAAAAGCT 237

Qy   1685 tcactcatttggtcactgtcctggagcgatatggccaagtattttcaaagctttgtcctg 1744
          | ||||||||| | |||||| ||||| ||||||||||||  || ||||| | ||| |||
Db    238 TTACTCATTTGATTACTGTCTTGGAGAGATATGGCCAAGTCATTGCAAAGATGTGTACTG 297

Qy   1745 ataacgataagcaggtgatgctattatctcaagtgagtacatactggaaaaacaatgtac 1804
          || | ||| | || || | |||  |  |  |||| ||| | |||||| |||||| ||  |
Db    298 ATGATGATCAACAAGTTAAGCTGATTACGGAAGTAAGTTCTTACTGGCAAAACAGTGCGC 357

Qy   1805 aaatgacggcggtggcaattgataggatgatgggttatagactagtatctaatcaggcaa 1864
          |||||||  ||  | || || ||  |||||||| ||||| | ||  |||||||  ||| |
Db    358 AAATGACAGCTATTGCCATCGACCGGATGATGAGTTATCGGCTTATATCTAATTTGGCCA 417

Qy   1865 ttgttagatgggtgttctctccagaaaatgttgatcagtttcatgtgtctgatcagccat 1924
          | || |||||||| ||||||||   ||| ||||| |||||||| |||||     | |
Db    418 TAGTGAGATGGGTTTTCTCTCCACTGAATCTTGATCGGTTTCATGTTTCTGACAGTTCGT 477

Qy   1925 gggagatacttggcaatgctcttaacaagacttataaccgtatctctgatttgaggaaag 1984
          |||||||  ||  | ||||||  | |  ||||| ||||| || || ||||||  |||| ||||
Db    478 GGGAGATCCTCAGGAATGCTGTCAGTAAGACGTATAATCGGATTTCTGATCTGAGAAAAG 537

Qy   1985 atatatcaaacattacgaaaaatgttttggttgctgagaaagcttcagccaatgcacgag 2044
          | || |||  | ||  ||| | |||| |  | |||||  |||||| || | |  ||  |||
Db    538 AGATTTCATCCCTTGAGAAGAGTGTTGTCTTGGCTGA-AAAGCTGCATCTAGAGCCAGAG 596

Qy   2045 tagagttggaggctgctgagagcaaactttccctagtggaaggtgaacccgttcttggtg 2104
           ||||||  ||  |||| |||   || ||      |  | || ||  ||||| ||||||||
Db    597 AAGAGTTAGAATCTGCCGAGTCTAAGCTGAGTGTCATAGATGGGGAACCTGTTCTTGGGT 656

Qy   2105 agaatcc 2111
            || ||
Db    657 GAAACCC 663


RESULT  12
BI935968
LOCUS       BI935968                643 bp    mRNA    linear   EST 18-OCT-2001
DEFINITION  EST555857 tomato flower, anthesis Lycopersicon esculentum cDNA
            clone cTOD24H3 5' end, mRNA sequence.
ACCESSION   BI935968
VERSION     BI935968.1  GI:16250440
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 643)
  AUTHORS   van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
            Utterback,T., Van Aken,S., Ronning,C.M., Nierman,W., Fraser,C.M.,
            Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
  TITLE     Generation of ESTs from tomato flower tissue, anthesis (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            This clone is available through the Clemson University Genomics
            Institute
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .643
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cTOD24H3"
                     /clone_lib="tomato flower, anthesis"
                     /tissue_type="flower"
                     /dev_stage="anthesis"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; supplier: Cornell University; sequencing: The
                     Institute for Genomic Research; Flower buds and flowers
                     were taken from greenhouse plants (4-8 wks old, TA496).
                     They were immediately frozen in liquid nitrogen and then
                     size-separated while remaining frozen."
BASE COUNT      183 a    114 c    163 g    183 t
ORIGIN


  Query Match           10.8%;  Score 294.2;  DB 10;  Length 643;
  Best Local Similarity  68.0%;  Pred. No. 2.8e-46;
  Matches  440;  Conservative    0;  Mismatches  203;  Indels    4;  Gaps    2;

Qy   1445 ctgccctagaagaattacttcctccaaaagctggtccgaattttatgtattccttggaag 1504
          ||||  | || ||  ||||||| |||| || ||| ||  | ||||  |||     |||||
Db      1 CTGCTTTGGAGGAGCTACTTCCACCAAGAGGTGGACCACAATTTAAATATAGTGCGGAAG 60

Qy   1505 aaggtaaagagaaaacagaagaacagcaattgtcagccgaattgagcaggaaggtcaagg 1564
          | ||||  ||   ||| ||    |    | | ||    ||| | |     | ||| || |
Db     61 ATGGTACTGATCCAACTGAG---CGTGCACTTTCTTTAGAACTCAAGGACATGGTGAAAG 117

Qy   1565 aaaaacaaaccgcacgtgacatgatagtgtggattgaagaaacgatatatccagttcatg 1624
            |   |||| || || || |||||    ||| | |||||||   | | |||||  ||||
Db    118 GTAGGAAAACTGCGCGGGAAATGATCTCCTGGGTAGAAGAAAATGTTTTTCCAGCCCATG 177

Qy   1625 gttttgaagttactcttacaatagttgtacagaccttacttgacatcggatcaaaaagtt 1684
          | |||||  |||| |||  | | ||||| || ||  | || ||||||||||||||||| |
Db    178 GGTTTGATATTACCCTTGGAGTTGTTGTTCAAACACTTCTCGACATCGGATCAAAAAGCT 237

Qy   1685 tcactcatttggtcactgtcctggagcgatatggccaagtattttcaaagctttgtcctg 1744
          | ||||||||| | |||||| ||||| ||||||||||||  || ||||| | ||| |||
Db    238 TTACTCATTTGATTACTGTCTTGGAGAGATATGGCCAAGTCATTGCAAAGATGTGTACTG 297

Qy   1745 ataacgataagcaggtgatgctattatctcaagtgagtacatactggaaaaacaatgtac 1804
          || | ||| | || || | |||  |  |  |||| ||| | |||||| |||||| ||  |
Db    298 ATGATGATCAACAAGTTAAGCTGATTACGGAAGTAAGTTCTTACTGGCAAAACAGTGCGC 357

Qy   1805 aaatgacggcggtggcaattgataggatgatgggttatagactagtatctaatcaggcaa 1864
          |||||||  ||  | || || ||  |||||||| ||||| | ||  |||||||  ||| |
Db    358 AAATGACAGCTATTGCCATCGACCGGATGATGAGTTATCGGCTTATATCTAATTTGGCCA 417

Qy   1865 ttgttagatgggtgttctctccagaaaatgttgatcagtttcatgtgtctgatcagccat 1924
          | || |||||||| ||||||||   ||| ||||| |||||||| |||||     | |
Db    418 TAGTGAGATGGGTTTTCTCTCCACTGAATCTTGATCGGTTTCATGTTTCTGACAGTTCGT 477

Qy   1925 gggagatacttggcaatgctcttaacaagacttataaccgtatctctgatttgaggaaag 1984
          |||||||  ||  | ||||||  | |  ||||| ||||| || || ||||||  |||| ||||
Db    478 GGGAGATCCTCAGGAATGCTGTCAGTAAGACGTATAATCGGATTTCTGATCTGAGAAAAG 537

Qy   1985 atatatcaaacattacgaaaaatgttttggttgctgagaaagcttcagccaatgcacgag 2044
          | || |||  | ||  || |||   | |  | |||||  |||||| || | |  ||  |||
Db    538 AGATTTCATCCCTTGAGAGAAAGTGTGTCTTGGCTGA-AAAGCTGCATCTAGAGCCAGAG 596

Qy   2045 tagagttggaggctgctgagagcaaactttccctagtggaaggtgaa 2091
```

```
          |||||| ||  |||| |||   || ||     |  | || || | |
Db    597 AAGAGTTAGAATCTGCCGAGTCTAAGCTGAGTGTCATAGATGGGGGA 643


RESULT 13
AW267982
LOCUS       AW267982                637 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  EST306204 DSIR Medicago truncatula cDNA clone pDSIR-8H13, mRNA
            sequence.
ACCESSION   AW267982
VERSION     AW267982.1  GI:6654882
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 637)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Peng
            ,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S.,
            Holt,I.E. and Fraser,C.M.
  TITLE     ESTs from roots of Medicago truncatula after inoculation with
            Phytophthora medicaginis
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            Minnesota EST name:M250435e
            TIGR sequence name:MTBAG43TK
            More information, including clone ordering, is available at. .
            http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .637
                     /organism="Medicago truncatula"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pDSIR-8H13"
                     /clone_lib="DSIR"
                     /tissue_type="roots infected with Phytophthora
                     medicaginis"
                     /dev_stage="roots harvested at 10 days post inoculation
                     with Phytophthora medicaginis"
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from
                     roots harvested at 10 days post inoculation with
                     Phytophthora medicaginis. The cDNA was directionally
                     ligated into the Uni-ZAP XR vector from Stratagene and
                     packaged using Gigapack III Gold packaging extracts.
                     Plasmids containing cDNA inserts were excised from the
                     recombinant lambda-ZAP phage using Ex-Assist helper phage
                     and propagated in XLOLR cells. Note: EST may be of fungal
                     origin."
BASE COUNT      173 a    131 c    138 g    195 t
ORIGIN


  Query Match           10.6%;  Score 288.2;  DB 9;  Length 637;
  Best Local Similarity   71.3%;  Pred. No. 3.9e-45;
  Matches  380;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy      23 cctctcggaagaagatgagcaattggaaaactcttctccttcgcatcggcgaaaagggac 82
           | |||  |   | |||||    |||||| ||  ||  | || || ||||| || ||| | |
Db     105 CTTCTTCGTCAACGATGAATGGTTGGAGAAGCCTCATTCTACGGATCGGTGATAAGAGTC 164

Qy      83 ctgagtacggcacttcctccgactacaaagaccacatcgagacttgtttcggtgtcattc 142
           | || || || |||||||||||||| |||||| ||||| || || ||||| |||||| |||
```

```
Db    165 CCGAATATGGAACTTCCTCCGACTTCAAAGATCACATTGAAACCTGTTTTGGTGTCCTTC 224

Qy    143 gtagagaaaatcgagcgttctggagatcaagttttgccttttctactacaatgtgctgaac 202
           || | ||  ||||    |||  |  | |  |||||   || || ||   |||||||||| |
Db    225 GTCGTGAGCTCGACAACTCTCAAACTGAGATTTTGGAGTTCCTTCTTACATGTGCTGAGC 284

Qy    203 aattgcctcataagattccttttgtatgggactttgattggtttgttgaacttggagaatg 262
           |||||||||| |||||||||| ||||| ||  |||||||  || | ||||||||  ||||
Db    285 AATTGCCTCACAAGATTCCTTTCTATGGAACCCTGATTGGCCTGATAAACTTGGAAAATG 344

Qy    263 aagattttgtccagaagctagtagaaagtgtccacgctaatttccaggtcgctttagatt 322
           | || |||||  | |  ||||| ||        |   |||| || ||||  ||| |||||
Db    345 AGGACTTTGTGAAAAGACTAGTGGAGCAAACTCGATCTAAATTTCAGGATGCTCTAGATA 404

Qy    323 ctggcaactgcaacagtatccgtatattgcttcgctttatgacttccctgttgtgcagta 382
           |||| ||||||||  |  | ||||| ||  | ||  | ||||||    || |||||||||
Db    405 CTGGAAACTGCAATGGAGTTCGTATTTTAATGCGGCTGATGACTGTTATGATGTGCAGTA 464

Qy    383 aggttattcaacctgcttctttgattgtcgtcttcgaaacattgctatcatctgctgcca 442
           | | | |||||||   |||| |  | |  ||||| |||| |||  ||||||| |||||||
Db    465 AAGCTCTTCAACCAAATTCTCTAGTGGATGTCTTTGAAATATTTTTATCATCAGCTGCCA 524

Qy    443 ctactgtggatgaagagaaaggaaatccatcatggcagccacaagctgactttacgtta 502
           | || || ||||| || ||||| ||||| | ||||||||||   |||||||||||| | |
Db    525 CAACAGTAGATGAGGAAAAAGGGAATCCTTTATGGCAGCCATGCGCTGACTTTTACATAA 584

Qy    503 tatgcatcttgtccagcctcccgtggggaggatcagaactcgctgagcaagtt 555
            ||||| || ||  ||||||| ||||| ||  | ||||| | ||||||||||
Db    585 CTTGCATTTTATCATGCCTCCCATGGGGTGGGGCTGAACTTGTTGAGCAAGTT 637


RESULT 14
AL365935
LOCUS       AL365935                466 bp    mRNA    linear   EST 03-AUG-2000
DEFINITION  MtBA03C07F1 MtBA Medicago truncatula cDNA clone MtBA03C07 T3, mRNA
            sequence.
ACCESSION   AL365935
VERSION     AL365935.1  GI:9665688
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from nitrogen-starved roots
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBA03C07"
                     /clone_lib="MtBA"
                     /tissue_type="root tips"
                     /dev_stage="harvested after 3 days of N-starvation"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber for 14
                     days on nitrogen-rich medium followed by 3 days on N-free
                     medium. RNA was extracted from root tips (1-3 cm). cDNA
```

was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zapXR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."

BASE COUNT 132 a 82 c 106 g 146 t
ORIGIN

Query Match 10.1%; Score 273.2; DB 9; Length 466;
Best Local Similarity 74.5%; Pred. No. 3.1e-42;
Matches 344; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
Qy    192 atgtgctgaacaattgcctcataagattcctttgtatgggactttgattggtttgttgaa 251
          ||||||||| |||||||||| |||||||||| ||||| ||  |||||||  || | ||
Db      4 ATGTGCTGAGCAATTGCCTCACAAGATTCCTTTCTATGGAACCCTGATTGGCCTGATAAA 63

Qy    252 cttggagaatgaagattttgtccagaagctagtagaaagtgtccacgctaatttccaggt 311
          |||||| ||||| || |||||  | |  ||||| ||       |   |||| || ||||
Db     64 CTTGGAAAATGAGGACTTTGTGAAAAGACTAGTGGAGCAAACTCGATCTAAATTTCAGGA 123

Qy    312 cgctttagattctggcaactgcaacagtatccgtatattgcttcgctttatgacttccct 371
           ||| ||||| |||| |||||||||  |  | ||||| ||  | ||  | ||||||    |
Db    124 TGCTCTAGATACTGGAAACTGCAATGGAGTTCGTATTTTAATGCGGCTGATGACTGTTAT 183

Qy    372 gttgtgcagtaaggttattcaacctgcttctttgattgtcgtcttcgaaacattgctatc 431
          | |||||||||| | | |||||||   |||| |  | |  ||||| |||| |||  ||||
Db    184 GATGTGCAGTAAAGCTCTTCAACCAAATTCTCTAGTGGATGTCTTTGAAATATTTTTATC 243

Qy    432 atctgctgccactactgtggatgaagagaaaggaaatccatcatggcagccacaagctga 491
          ||| |||||||| || || ||||| || ||||| ||||| | ||||||||||   |||||
Db    244 ATCAGCTGCCACAACAGTAGATGAGGAAAAAGGGAATCCTTTATGGCAGCCATGCGCTGA 303

Qy    492 cttttacgttatatgcatcttgtccagcctcccgtggggaggatcagaactcgctgagca 551
          ||||||| | |  ||||| || ||  ||||||| ||||| ||  | ||||| | ||||||
Db    304 CTTTTACATAACTTGCATTTTATCATGCCTCCCATGGGGTGGGGCTGAACTTGTTGAGCA 363

Qy    552 agttcctgatgagattgaaagagtgttagttgggatacaagcttatttgagcatccgaaa 611
          |||||| || || ||||| |||||  | ||||| ||| ||||||||||||||||||||||
Db    364 AGTTCCCGAAGATATTGAGAGAGTAATGGTTGGTATAGAAGCTTATTTGAGCATCCGAAA 423

Qy    612 gaattcttccacctctgggttaaacttttttcacaacggaga 653
          | || ||||   |||||| |||  ||||||| | || |  ||
Db    424 GCATACTTCTGACTCTGGTTTATCCTTTTTTGAGAATGATGA 465
```

RESULT 15
BE590437
LOCUS BE590437 632 bp mRNA linear EST 17-AUG-2000
DEFINITION SB177 Sugar Beet germination cDNA library Beta vulgaris cDNA clone ys016f11 5', mRNA sequence.
ACCESSION BE590437
VERSION BE590437.1 GI:9843476
KEYWORDS EST.
SOURCE Beta vulgaris.
ORGANISM Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae; Caryophyllales; Chenopodiaceae; Beta.
REFERENCE 1 (bases 1 to 632)
AUTHORS de los Reyes,B.G., McGrath,J.M., Myers,S. and Derrico,C.
TITLE Differential gene expression in sugar beet seedlings (Beta vulgaris) germinated under stress conditions
JOURNAL Unpublished (2000)
COMMENT Contact: J. Mitchell McGrath
Sugar Beet Genetics Lab., USDA/ARS/Sugar Beet and Bean Research Unit
Michigan State University
494 Plant and Soil Sci. Bldg., East Lansing MI 48824, USA
Tel: (517)-432-2355
Fax: (517)-337-6782
Email: mitchmcg@pilot.msu.edu
Seq primer: T3
High quality sequence stop: 624.

FEATURES Location/Qualifiers
source 1. .632
/organism="Beta vulgaris"
/cultivar="USH20"
/db_xref="taxon:161934"
/clone="ys016f11"
/clone_lib="Sugar Beet germination cDNA library"
/tissue_type="whole seedlings"
/dev_stage="4-day germination under stress (salt/NaCl, dehydration/Mannitol and anaerobic stress)"
/lab_host="SOLR"
/note="Organ: seeds; Vector: pBluescript II KS (+) excised from Lambda Uni-Zap XR; Site_1: EcoRI; Site_2: XhoI; cDNAs were derived from reverse transcription of pooled mRNA samples from stressed seedlings (germinated for 4 days in 150mM NaCl, 200mM Mannitol, submerged in distilled water and 0.3% Hydrogen peroxide). The cDNA library was generated by directional ligation of the cDNAs in the EcoRI and XhoI sites of Lambda Uni-Zap XR vector (Stratagene). The library was excised as a pBluescript plasmid from Lambda Uni-Zap XR by ExAssist helper phage and plated on SOLR host cells. "

BASE COUNT 162 a 150 c 131 g 189 t
ORIGIN

Query Match 9.9%; Score 267.8; DB 10; Length 632;
Best Local Similarity 66.6%; Pred. No. 3.1e-41;
Matches 399; Conservative 0; Mismatches 197; Indels 3; Gaps 1;

```
Qy     37 atgagcaattggaaaactcttctccttcgcatcggcgaaaagggacctgagtacggcact 96
          ||||| |  |||||||| ||  | || || ||||||||||  | || || ||||||
Db     30 ATGAGTAGCTGGAAAACCCTAATTCTCCGAATCGGCGAAAAATGTCCCGAATACGGCGGA 89

Qy     97 tcctccgactacaaagaccacatcgagacttgtttcggtgtcattcgtagagaaatcgag 156
            |  |||| |||||| |||||||| | ||| | ||   |  || |  | ||| | ||
Db     90 AACCTAGACTTCAAAGAACACATCGACAGTTGCTACGCCATTCTTTGGCGCGAATTAGAC 149

Qy    157 cgttctggagatcaagttttgccttt---tctactacaatgtgctgaacaattgcctcat 213
          |      | ||| || |    | |  ||      | || ||||| || ||||||||||||||
Db    150 CAACAAGAAGACCACGAAATCCTATTCTACGTCCTCCAATGCGCCGAACAATTGCCTCAC 209

Qy    214 aagattcctttgtatgggactttgattggtttgttgaacttggagaatgaagattttgtc 273
          ||   ||| || || || || || || ||| |  | ||| | ||||| ||||||  | | |
Db    210 AAATCTCCCTTTTACGGCACCTTAATCGGTCTCCTAAACCTCGAGAAAGAAGATCTCGGC 269

Qy    274 cagaagctagtagaaagtgtccacgctaatttccaggtcgctttagattctggcaactgc 333
          |     |||| ||||     || | ||| | ||    ||  |  | || |  || |||
Db    270 AAATTAGTAGTCGAAACCACTCATTCCAATCTTCAACAAGCAATTCAGTCGGAAAATTGC 329

Qy    334 aacagtatccgtatattgcttcgctttatgacttccctgttgtgcagtaaggttattcaa 393
          |||| || ||| | ||| | || ||| | |||     || |||| || |||||| |||||
Db    330 GACAGAATTCGTGTTTTGTTGCGATTTCTCACTGTTTTGATGTGTAGCAAGGTTCTTCAA 389

Qy    394 cctgcttctttgattgtcgtcttcgaaacattgctatcatctgctgccactactgtggat 453
          |||||||||||| |||| || || || || ||||| || || ||||| || | ||| |||
Db    390 CCTGCTTCTTTGGTTGTTGTGTTTGAGACTTTGCTTTCGTCAGCTGCTACCATTGTTGAT 449

Qy    454 gaagagaaaggaaatccatcatggcagccacaagctgacttttacgttatatgcatcttg 513
          || |||||||| |||||  | |||||  | |  ||||| |||||  | |  ||  | |||
Db    450 GAGGAGAAAGGCAATCCTGCTTGGCAAGCGCGCGCTGATTTTTATATAAGTTGTGTTTTG 509

Qy    514 tccagcctcccgtggggaggatcagaactcgctgagcaagttcctgatgagattgaaaga 573
          ||  |||| |||||||||||  | || ||    |||||| ||||||| || |||||  ||
Db    510 TCATGCCTTCCGTGGGGAGGTGCTGAGCTTATGGAGCAACTTCCTGAGGATATTGATCGA 569

Qy    574 gtgttagttgggatacaagcttatttgagcatccgaaagaattcttccacctctgggtt 632
          | || |||||||| |||||||||| || || || |   |  ||      |||||| |
Db    570 CTTTTGGTTGGGATTCAAGCTTATTTCAGTATTCGGAGACACCCTATTGATTCTGGGCT 628
```

Search completed: August 20, 2002, 12:28:33
Job time: 3192 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 11:39:56 ; Search time 105.98 Seconds
(without alignments)
6294.967 Million cell updates/sec

Title: US-09-882-986-1
Perfect score: 2716
Sequence: 1 aaagagacgaactgaagaaa.....ctttaaaaaaaaaaaaaaaa 2716

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	45.2	1.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	45	1.7	1386	2	US-08-897-340-1	Sequence 1, Appli
	3	45	1.7	1386	3	US-09-252-329-1	Sequence 1, Appli
	4	42.8	1.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c	5	42.4	1.6	276	4	US-09-328-111-820	Sequence 820, App
	6	42.2	1.6	1248	4	US-09-020-956-171	Sequence 171, App
	7	42.2	1.6	1248	4	US-09-030-607-171	Sequence 171, App
	8	42.2	1.6	1248	4	US-09-439-313-171	Sequence 171, App
	9	40.8	1.5	12720	1	US-08-403-866-11	Sequence 11, Appl
c	10	40.2	1.5	3138	1	US-07-867-106-4	Sequence 4, Appli
	11	40	1.5	1214	4	US-09-439-313-225	Sequence 225, App
	12	40	1.5	1215	4	US-09-439-313-326	Sequence 326, App
	13	39.2	1.4	642	3	US-08-946-026-33	Sequence 33, Appl
c	14	39	1.4	4500	2	US-08-743-637B-35	Sequence 35, Appl
c	15	39	1.4	4500	3	US-08-526-840B-35	Sequence 35, Appl
c	16	39	1.4	5000	3	US-09-104-070-3	Sequence 3, Appli
c	17	38.8	1.4	5852	1	US-07-867-106-2	Sequence 2, Appli
	18	38.8	1.4	15766	4	US-09-338-907-73	Sequence 73, Appl
	19	38.8	1.4	15766	4	US-09-218-207-73	Sequence 73, Appl
	20	38.8	1.4	37950	4	US-09-338-907-183	Sequence 183, App
	21	38.8	1.4	37950	4	US-09-218-207-183	Sequence 183, App
	22	38	1.4	1751	2	US-08-955-138-3	Sequence 3, Appli
	23	37.4	1.4	358	2	US-08-721-488-9	Sequence 9, Appli
	24	37.4	1.4	746	4	US-09-013-810-1	Sequence 1, Appli
	25	37.2	1.4	1098	3	US-09-248-335-35	Sequence 35, Appl
	26	36.8	1.4	547	1	US-08-131-365B-41	Sequence 41, Appl
	27	36.8	1.4	547	2	US-08-668-123-41	Sequence 41, Appl
	28	36.8	1.4	11517	1	US-07-920-281C-1	Sequence 1, Appli
	29	36.8	1.4	11517	4	US-08-466-277-1	Sequence 1, Appli
	30	36.4	1.3	1117	4	US-09-247-373B-33	Sequence 33, Appl
	31	36.4	1.3	2407	4	US-09-370-807-7	Sequence 7, Appli
c	32	36.4	1.3	4450	3	US-08-617-860B-2	Sequence 2, Appli
	33	36.2	1.3	2550	6	5258287-23	Patent No. 5258287
	34	35.8	1.3	958	2	US-08-757-046A-5	Sequence 5, Appli
	35	35.8	1.3	958	3	US-09-447-208-5	Sequence 5, Appli
	36	35.8	1.3	958	3	US-09-135-988-5	Sequence 5, Appli
	37	35.8	1.3	958	4	US-09-277-716-5	Sequence 5, Appli
	38	35.8	1.3	958	4	US-08-597-274A-5	Sequence 5, Appli
	39	35.8	1.3	2852	3	US-09-027-137-2	Sequence 2, Appli
	40	35.8	1.3	4880	3	US-09-031-563-1	Sequence 1, Appli
	41	35.8	1.3	5125	3	US-09-031-563-4	Sequence 4, Appli
	42	35.6	1.3	1166	5	PCT-US96-12129B-1	Sequence 1, Appli
	43	35.6	1.3	1582	3	US-08-545-196B-10	Sequence 10, Appl
	44	35.6	1.3	1582	3	US-08-545-196B-12	Sequence 12, Appl
	45	35.6	1.3	3214	1	US-08-484-105-17	Sequence 17, Appl

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14
```

```
  Query Match             1.7%;  Score 45.2;  DB 1;  Length 7218;
  Best Local Similarity   5.6%;  Pred. No. 0.0041;
  Matches   23;  Conservative  213;  Mismatches  176;  Indels    0;  Gaps    0;

Qy    2046 agagttggaggctgctgagagcaaacttttccctagtggaaggtgaacccgttcttggtga 2105
           |||| | || |    ||  |  :::        :: :::::: :::   :    :: ::
Db    1455 AGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1396

Qy    2106 gaatccagcgaagatgaagcgtttaaaatcaacagtggagaagacaggggaagcggagtt 2165
           :::   :: ::::: :::: :   ::::  :: :: :::::::: :::::::: ::::
Db    1395 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1336

Qy    2166 atctcttcgggagtccctagaggcaaaagaggctcttcttaacagagctctctctgagac 2225
           :       :::::     ::::: :::::::        :: ::::        ::::
Db    1335 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1276

Qy    2226 cgaggttttactgctcttgctgttccaaagtttcttaggtgtactgaaggaacggctccc 2285
            ::::    :  :     :  :    ::::      ::: : :  :::::::: ::
Db    1275 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1216

Qy    2286 agatccaactaaagtgagatcagtgcaggatctaaaatctataggtgctgaagatgacaa 2345
           :::   ::  :::: ::::  :: : ::::   ::::   : ::: :  ::::: :: ::
Db    1215 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1156

Qy    2346 gccatctgcgatggacgtggacagcgagaatggaaacccaaagaagagttgcgaagtcgg 2405
           :  :   : :: ::: : ::: :: ::::: :::::   :::::::::  : ::::  ::
Db    1155 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1096

Qy    2406 tgagagagaacagtggtgcttatcaacacttggctatctcacggcatttaca 2457
           ::::::::: :: :: :   :  :: :  | || | ||| |   |  | ||
Db    1095 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACCTGCA 1044


RESULT   2
US-08-897-340-1
; Sequence 1, Application US/08897340
; Patent No. 5955306
;  GENERAL INFORMATION:
;    APPLICANT:  Gimeno, Carlos J. and Errada, Patrick, R.
;    TITLE OF INVENTION:  Weight Control Pathway Genes and Uses
;    TITLE OF INVENTION:  Therefor
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/897,340
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/715,032
;      FILING DATE:  17-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silveri, Jean M.
;      REGISTRATION NUMBER:  39,030
;      REFERENCE/DOCKET NUMBER:  MNI-005CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1386 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-897-340-1


  Query Match             1.7%;  Score 45;  DB 2;  Length 1386;
  Best Local Similarity   55.1%;  Pred. No. 0.0016;
  Matches   81;  Conservative    2;  Mismatches   64;  Indels    0;  Gaps    0;

Qy    2570 tcccattacattaatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgt 2629
           ||||  ||||     |  |||  |   |  |  | | |||||  |   || ||||||  |
Db    1219 TCCCTGTACANATTGCCNCCTTGTGGCACGTTGACCCAACCTTACCAGTTGGTTTTTCAT 1278

Qy    2630 tatgagattctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaaga 2689
           | |  |   ||    |  | ||| ||  || | |     |: :|| | ||||||| || |
Db    1279 TTTTTGTCCCTTTCCCCTAGATCCAGAAATAAAGTTTAAGRGRAGSGCCAAAAAAAAAAA 1338

Qy    2690 gtttaaactttaaaaaaaaaaaaaaa 2716
           |||    ||||||||||||||||
Db    1339 AAAAAAAAAAAAAAAAAAAAAAAAA 1365


RESULT   3
US-09-252-329-1
; Sequence 1, Application US/09252329
; Patent No. 6147192
;  GENERAL INFORMATION:
;    APPLICANT:  Gimeno, Carlos J. and Errada, Patrick, R.
;    TITLE OF INVENTION:  Weight Control Pathway Genes and Uses
;    TITLE OF INVENTION:  Therefor
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/252,329
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/897,340
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silveri, Jean M.
;      REGISTRATION NUMBER:  39,030
;      REFERENCE/DOCKET NUMBER:  MNI-005CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1386 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-252-329-1


  Query Match             1.7%;  Score 45;  DB 3;  Length 1386;
  Best Local Similarity   55.1%;  Pred. No. 0.0016;
  Matches   81;  Conservative    2;  Mismatches   64;  Indels    0;  Gaps    0;

Qy    2570 tcccattacattaatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgt 2629
           ||||  ||||     |  |||  |   |  |  | | |||||  |   || ||||||  |
```

```
Db   1219 TCCCTGTACANATTGCCNCCTTGTGGCACGTTGACCCAACCTTACCAGTTGGTTTTTCAT 1278

Qy   2630 tatgagattctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaaga 2689
          | |  |   ||    |  | ||| ||  || | |     |: :|| | ||||||| || |
Db   1279 TTTTTGTCCCTTTCCCCTAGATCCAGAAATAAAGTTTAAGRGRAGSGCCAAAAAAAAAAA 1338

Qy   2690 gtttaaactttaaaaaaaaaaaaaaa 2716
             |||    ||||||||||||||||
Db   1339 AAAAAAAAAAAAAAAAAAAAAAAAAAA 1365


RESULT   4
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             1.6%;  Score 42.8;  DB 1;  Length 7218;
  Best Local Similarity   3.8%;  Pred. No. 0.022;
  Matches   14;  Conservative  202;  Mismatches  154;  Indels    0;  Gaps    0;

Qy   1006 ttgaatggatgtcggaaggagtgtgcatcctacatggctaatcttcctgttacatttcgg 1065
          |||  || | ||||    |   |    ||::: : :  ::  :::::::: :: : ::::
Db   1039 TTGGCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1098

Qy   1066 tacgagtatcttatggcagagacactattttctcagatactgctgctaccccagccacca 1125
          : :   : :::: :  :     : :: :::::::   : :: :: :: :::: :: ::
Db   1099 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1158

Qy   1126 ttcaagactctttattatacactcgtgattatggatctttgtaaggctcttccgggtgcc 1185
          :::    :::::: :: : : ::: :  :: :  :::: :    ::::::  :: ::
Db   1159 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1218

Qy   1186 tttcctgctgttgttgctggcgctgttcgtgcactatttgagaaaatatccgacttagac 1245
          :::::: :: :: :: ::  : :: ::: : : :: :::       : :::  :::   :
Db   1219 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1278

Qy   1246 atggaatccaggacgcgtcttatcctctggttttctcaccacttatccaacttccaattc 1305
          :    :::    : : :::: :::::: :::::: :: ::: :::  :::::  :::
Db   1279 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1338

Qy   1306 atctggccgtgggaagagtgggcttttgtgttggatcttcccaagtgggcccctaagcgt 1365
          :::  :: :      :  ::::: : ::  :::::::  :  :::::   : : :
Db   1339 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1398

Qy   1366 gtatttgttc 1375
          : ::: :::
Db   1399 YYYYYYYYYY 1408


RESULT   5
US-09-328-111-820/c
; Sequence 820, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
;  APPLICANT: Endege, Wilson O.
;  APPLICANT: Steinmann, Kathleen E.
;  APPLICANT: Astle, Jon H.
;  APPLICANT: Burgess, Christopher C.
;  APPLICANT: Bushnell, Steven E.
;  APPLICANT: Carroll III, Eddie
;  APPLICANT: Catino, Theodore J.
;  APPLICANT: Derti, Adnan
;  APPLICANT: Ford, Donna M.
;  APPLICANT: Lewis, Marcia E.
;  APPLICANT: Monahan, John E.
;  APPLICANT: Schlegel, Robert
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS
;  FILE REFERENCE: CCD-257 (US)
;  CURRENT APPLICATION NUMBER: US/09/328,111
;  CURRENT FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: US 60/088,801
;  EARLIER FILING DATE: 1998-06-10
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 820
;   LENGTH: 276
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-328-111-820


  Query Match             1.6%;  Score 42.4;  DB 4;  Length 276;
  Best Local Similarity   48.7%;  Pred. No. 0.0034;
  Matches  115;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy    194 gtgctgaacaattgcctcataagattcctttgtatgggactttgattggtttgttgaact 253
          ||||    | ||| ||| | ||| |  |  | |||   || ||  ||||  |  ||||
Db    259 GTGCACGCCTATTACCTGAGAAGCTGACAATTTATACAACATTAGTTGGACTACTGAATG 200

Qy    254 tggagaatgaagattttgtccagaagctagtagaaagtgtccacgctaatttccaggtcg 313
              |||| |  ||||||      |  | ||||||    |      | |  |  | |
Db    199 CCAGGAATTACAATTTTGGTGGAGAATTTGTAGAAGCCATGATTCGTCAACTTAAAGAAT 140

Qy    314 ctttagattctggcaactgcaacagtatccgtatattgcttcgctttatgacttccctgt 373
          | ||  |  |   ||| |  ||       |      ||| | || ||| |  ||   ||
Db    139 CATTGAAAGCAAACAATTATAATGAAGCCGTGTATTTGGTCCGTTTTTTATCTGATCTTG 80

Qy    374 tgtgcagtaaggttattcaacctgcttctttgattgtcgtcttcgaaacattgcta 429
          ||    || | || |||    |  | ||  || ||| |  || |||| ||  ||
Db     79 TGAATTGTCATGTGATTGCCGCCCCATCAATGGTTGCTATGTTTGAAAATTTTGTA 24
```

```
RESULT   6
US-09-020-956-171
; Sequence 171, Application US/09020956
; Patent No. 6261562
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillin, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  178
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/020,956
;      FILING DATE:  09-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  171:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1248 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-020-956-171


  Query Match             1.6%;  Score 42.2;  DB 4;  Length 1248;
  Best Local Similarity   55.2%;  Pred. No. 0.011;
  Matches   80;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy   2570 tcccattacattaatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgt 2629
          ||||  ||||     |  |||  |  |  |  | | |||||  |   || ||||||  |
Db   1104 TCCCTGTACACAGTGCCCCCTTGTGGCACGTTGACCCAACCTTACCAGTTGGTTTTTCAT 1163

Qy   2630 tatgagattctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaaga 2689
          | |  |   ||    |  | ||| ||  || | |     || ||| |  |||||| || |
Db   1164 TTTTNGTCCCTTTCCCCTAGATCCAGAAATAAAGTTTAAGAGAAGNGCAAAAAAAAAAAA 1223

Qy   2690 gtttaaactttaaaaaaaaaaaaa 2714
               |||    ||||||||||||||
Db   1224 AAAAAAAAAAAAAAAAAAAAAAAA 1248


RESULT   7
US-09-030-607-171
; Sequence 171, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  171:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1248 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-030-607-171


  Query Match             1.6%;  Score 42.2;  DB 4;  Length 1248;
  Best Local Similarity   55.2%;  Pred. No. 0.011;
  Matches   80;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy   2570 tcccattacattaatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgt 2629
          ||||  ||||     |  |||  |  |  |  | | |||||  |   || ||||||  |
Db   1104 TCCCTGTACACAGTGCCCCCTTGTGGCACGTTGACCCAACCTTACCAGTTGGTTTTTCAT 1163

Qy   2630 tatgagattctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaaga 2689
          | |  |   ||    |  | ||| ||  || | |     || ||| |  |||||| || |
Db   1164 TTTTNGTCCCTTTCCCCTAGATCCAGAAATAAAGTTTAAGAGAAGNGCAAAAAAAAAAAA 1223

Qy   2690 gtttaaactttaaaaaaaaaaaaa 2714
               |||    ||||||||||||||
Db   1224 AAAAAAAAAAAAAAAAAAAAAAAA 1248


RESULT   8
US-09-439-313-171
; Sequence 171, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 171
;   LENGTH: 1248
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1248)
;   OTHER INFORMATION: n = A,T,C or G
US-09-439-313-171


  Query Match             1.6%;  Score 42.2;  DB 4;  Length 1248;
  Best Local Similarity   55.2%;  Pred. No. 0.011;
  Matches   80;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy   2570 tcccattacattaatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgt 2629
          ||||  ||||     |  |||  |   |  |  | | |||||  |   || ||||||  |
Db   1104 tccctgtacacagtgccccttgtggcacgttgacccaaccttaccagttggtttttcat 1163

Qy   2630 tatgagattctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaaga 2689
          | |  |   ||    |  | ||| ||  || | |     || ||| |  |||||| || |
Db   1164 ttttngtccctttcccctagatccagaaataaagtttaagagaagngcaaaaaaaaaaa 1223

Qy   2690 gtttaaactttaaaaaaaaaaaaa 2714
               |||    ||||||||||||||
Db   1224 aaaaaaaaaaaaaaaaaaaaaaaa 1248


RESULT   9
US-08-403-866-11
; Sequence 11, Application US/08403866
; Patent No. 5643779
;  GENERAL INFORMATION:
;    APPLICANT:  Ehrlich, Stanislav
;    APPLICANT:  Godon, Jean-Jacques
;    APPLICANT:  Renault, Pierre
;    TITLE OF INVENTION:  Nucleic acid coding for an alpha-acetolactate
;    TITLE OF INVENTION:  synthase from Lactococcus and its applications
;    NUMBER OF SEQUENCES:  16
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,866
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30, 727
;      REFERENCE/DOCKET NUMBER:  20747/30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1600
;      TELEFAX:  (716) 263-1487
;      TELEX:  978450 (WUT)
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12720 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Lactococcus lactis subsp. lactis
US-08-403-866-11


  Query Match             1.5%;  Score 40.8;  DB 1;  Length 12720;
  Best Local Similarity   50.5%;  Pred. No. 0.13;
  Matches   99;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy   2521 gaagatgtgcatcctctctttctccaagccatatcttctgcacttcaattcccattacat 2580
          . ||  || |  |  |    |||||||      |   |  | || |  ||| |    ||  |
Db      9 GATAATCTTGAGTCATAATTTCTCCTTAATCTTATTAGTACATTAGAATCCATTATAATT 68

Qy   2581 taatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgttatgagattct 2640
          ||||| |  | | || | || || |||      |  ||| | | |||| | |  || ||
Db     69 TAATCATTTTATGTCTACCTAAAGCAACAAAATTGCTTGTATATTTTCTAACAAGCTTAA 128

Qy   2641 gattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaagagtttaaacttt 2700
               | ||   | || | | ||  ||| || |||    ||   |   |  |||| | |
Db    129 TTATGTGGATTTAATTGAATATTAAAGGGAGAAGTTGTAATCTATTTGTTGTTAAATTCT 188

Qy   2701 aaaaaaaaaaaaaaaa 2716
               || |.||| |||
Db    189 TGTTAATACAAATAAA 204


RESULT  10
US-07-867-106-4/c
; Sequence 4, Application US/07867106
; Patent No. 5389526
;  GENERAL INFORMATION:
;    APPLICANT:  Slade, Martin B
;    APPLICANT:  Chang, Andy C M
;    APPLICANT:  Williams, Keith L
;    TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
;    TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
;      STREET:  One Liberty Place 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/867,106
;      FILING DATE:  19920625
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PJ 7187
;      APPLICATION NUMBER:  PCT/AU90/00530
;      FILING DATE:  02-NOV-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feeney, Joanne Longo
;      REGISTRATION NUMBER:  35,134
;      REFERENCE/DOCKET NUMBER:  RICE-0002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3138 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
US-07-867-106-4


  Query Match             1.5%;  Score 40.2;  DB 1;  Length 3138;
  Best Local Similarity   55.3%;  Pred. No. 0.078;
  Matches   78;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy   2576 tacattaatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgttatgag 2635
          ||  ||    || |  ||| | | |  ||  ||  |   |  |||||||  |  || ||
```

```
Db   2083 TATGTTGGAATTTCATTTTTATTTTTTATTTAATTTAATTTATTTGTTTTAAGAAATAAG 2024

Qy   2636 attctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaagagtttaa 2695
          |     |     | |  |  || ||| | |||   ||| |   |||||| || |     ||
Db   2023 AAAAAAAAAAAAAAAAAATCTTTTTATGCAATCTGAAAAAAAAAAAAAAAAAAAAAAAAA 1964

Qy   2696 actttaaaaaaaaaaaaaaa 2716
          |    ||||||||||||||||
Db   1963 AAAAAAAAAAAAAAAAAAAA 1943


RESULT  11
US-09-439-313-225
; Sequence 225, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 225
;   LENGTH: 1214
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-439-313-225


  Query Match             1.5%;  Score 40;  DB 4;  Length 1214;
  Best Local Similarity   54.9%;  Pred. No. 0.049;
  Matches   79;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy   2570 tcccattacattaatcttcctctttcaatctcaatcaaacctgtctcttttgtttttttgt 2629
          ||||  ||||     |  |||  |   |  | | |||||  |   || ||||||  |
Db   1071 tccctgtacacagtgccccctgtggcacgttgacccaaccttaccagttggtttttcat 1130

Qy   2630 tatgagattctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaaga 2689
          | |  |   ||    |  | ||| ||  || | |     || ||| |  |||||| || |
Db   1131 tttttgtccctttcccctagatccagaaataaagtctaagagaagcgcaaaaaaaaaaaa 1190

Qy   2690 gtttaaactttaaaaaaaaaaaa 2713
               |||    ||||||||||||
Db   1191 aaaaaaaaaaaaaaaaaaaaaa 1214


RESULT  12
US-09-439-313-326
; Sequence 326, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 326
;   LENGTH: 1215
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-439-313-326


  Query Match             1.5%;  Score 40;  DB 4;  Length 1215;
  Best Local Similarity   54.9%;  Pred. No. 0.049;
  Matches   79;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy   2570 tcccattacattaatcttcctctttcaatctcaatcaaacctgtctcttttgtttttttgt 2629
          ||||  ||||     |  |||  |   |  | | |||||  |   || ||||||  |
Db   1072 tccctgtacacagtgccccctgtggcacgttgacccaaccttaccagttggtttttcat 1131

Qy   2630 tatgagattctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaaga 2689
          | |  |   ||    |  | ||| ||  || | |     || ||| |  |||||| || |
Db   1132 tttttgtccctttcccctagatccagaaataaagtctaagagaagcgcaaaaaaaaaaaa 1191

Qy   2690 gtttaaactttaaaaaaaaaaaa 2713
               |||    ||||||||||||
Db   1192 aaaaaaaaaaaaaaaaaaaaaa 1215


RESULT  13
US-08-946-026-33
; Sequence 33, Application US/08946026
; Patent No. 6034218
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Twardzik, Daniel R.
;    APPLICANT:  Mitcham, Jennifer L.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
;    TITLE OF INVENTION:  AND IMMUNODIAGNOSIS OF PROSTATE CANCER
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/946,026
;      FILING DATE:  07-OCT-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.424C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  642 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

US-08-946-026-33

Query Match 1.4%; Score 39.2; DB 3; Length 642;
Best Local Similarity 57.8%; Pred. No. 0.056;
Matches 89; Conservative 0; Mismatches 63; Indels 2; Gaps 1;

Qy 2565 tcaattcccattacattaatcttcctctttcaatctcaatcaaacctgtctcttttgttt 2624
Db 481 TGAAGTCACAAGATACATATATTTTTATTTTGACCTAAATTTGTACAGTCCCATTGTAAG 540

Qy 2625 tttgttatgagattctgattctgacatcaagtt--attaggaaattgaaaagagtcaaaa 2682
Db 541 TGTTGTTTCTAATTATAGATGTAAAATGAAATTTCATTTGTAATTGGAAAAAATCCAATA 600

Qy 2683 aacaagagtttaaactttaaaaaaaaaaaaaaa 2716
Db 601 AAAAGGATATTCATTTAAAAAAAAAAAAAAAAA 634

RESULT 14
US-08-743-637B-35/c
; Sequence 35, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414)277-5591
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4500 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
US-08-743-637B-35

Query Match 1.4%; Score 39; DB 2; Length 4500;
Best Local Similarity 53.6%; Pred. No. 0.23;
Matches 81; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 2522 aagatgtgcatcctctctttctccaagccatatcttctgcacttcaattcccattacatt 2581
Db 825 AAGTAGTTCAACTTGTCTTTTTCTCTGCTTTTTTGTCTTGAATGCCTTAAATATCATATT 766

Qy 2582 aatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgttatgagattctg 2641
Db 765 ATTCCCCTTTTCTAAGCTTCAATCCCTTCTCTCTATCTGTTATTTCGCATAGCTACTACT 706

Qy 2642 attctgacatcaagttattaggaaattgaaa 2672
Db 705 ATTTTTACAGCAATCTATCTTTAAATATACA 675

RESULT 15
US-08-526-840B-35/c
; Sequence 35, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENE
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4500 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
US-08-526-840B-35

Query Match 1.4%; Score 39; DB 3; Length 4500;
Best Local Similarity 53.6%; Pred. No. 0.23;
Matches 81; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 2522 aagatgtgcatcctctctttctccaagccatatcttctgcacttcaattcccattacatt 2581

```
Db     825 AAGTAGTTCAACTTGTCTTTTTCTCTGCTTTTTTGTCTTGAATGCCTTAAATATCATATT 766

Qy    2582 aatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgttatgagattctg 2641
           | ||  | | | | |   ||||||    || ||| | |  | ||| |    |  | |
Db     765 ATTCCCCTTTTCTAAGCTTCAATCCCTTCTCTCTATCTGTTATTTCGCATAGCTACTACT 706

Qy    2642 attctgacatcaagttattaggaaattgaaa 2672
           ||| | ||| |||  |||    ||||  | |
Db     705 ATTTTTACAGCAATCTATCTTTAAATATACA 675
```

Search completed: August 20, 2002, 13:51:46
Job time: 7910 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 11:41:01 ; Search time 413.38 Seconds
(without alignments)
11280.516 Million cell updates/sec

Title: US-09-882-986-1
Perfect score: 2716
Sequence: 1 aaagagacgaactgaagaaa.....ctttaaaaaaaaaaaaaaaa 2716

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	61.8	2.3	3224	23	ABL12159	Drosophila melanog
c	2	56.8	2.1	2902	23	ABL27678	Drosophila melanog
	3	50.8	1.9	2244	23	ABL13529	Drosophila melanog
	4	50.8	1.9	4420	23	ABL13528	Drosophila melanog
c	5	45.4	1.7	6242	24	ABL34148	Human immune syste
	6	45	1.7	1386	19	AAV11855	Homo sapiens Tub I
c	7	44.4	1.6	6552	22	AAS46678	Tumour suppressor
	8	44	1.6	6189	22	AAS46599	Tumour suppressor
	9	44	1.6	6189	24	ABL34592	Human metastasis a
c	10	43.6	1.6	669	22	AAL15089	Human breast cance
	11	43.4	1.6	2872	13	AAQ25449	Cholesterol 7-alph
	12	43	1.6	1456	22	AAD03657	Human serine prote
c	13	42.8	1.6	256	22	AAS07790	Cervical cancer pr
	14	42.8	1.6	612	22	AAH71471	Human cervical can
c	15	42.8	1.6	612	22	AAH71471	Human cervical can
c	16	42.6	1.6	396	22	AAF94862	Human ovarian canc
c	17	42.6	1.6	425	22	AAS60450	Human cancer agent
	18	42.6	1.6	464	22	AAH33241	Human colon cancer
	19	42.6	1.6	3084	22	AAS25910	Human cDNA encodin
c	20	42.4	1.6	268	22	AAH70080	Human cervical can
c	21	42.4	1.6	276	21	AAZ80736	Human colon cancer
	22	42.4	1.6	411	22	AAI88595	Human polynucleoti
	23	42.2	1.6	1248	19	AAV61249	cDNA sequence of p
	24	42.2	1.6	1248	19	AAV58644	Prostate tumour sp
	25	42.2	1.6	1248	21	AAA06407	Human immunogenic
	26	42.2	1.6	1248	22	AAS63615	Human prostate cDN
	27	42.2	1.6	1248	22	AAS10166	Human prostate tum
	28	42.2	1.6	1248	22	AAH93523	Human prostate-spe
	29	42.2	1.6	1248	22	AAH84837	Human prostate-spe
	30	42.2	1.6	1248	22	AAH02588	Prostate tumour an
c	31	42	1.5	165	22	AAI84674	Human polynucleoti
	32	41.8	1.5	198	22	AAS07727	Cervical cancer pr
	33	41.8	1.5	1767	22	AAS29126	cDNA encoding for
	34	41.6	1.5	176	22	AAH25484	Nucleotide fragmen
	35	41.6	1.5	398	22	AAI88388	Human polynucleoti
c	36	41.6	1.5	534	22	AAS60058	Human cancer agent
	37	41.6	1.5	2498	21	AAA27988	Wheat CCR4 transcr
	38	41.6	1.5	4091	22	AAI58457	Human polynucleoti
	39	41.4	1.5	230	22	AAH33118	Human colon cancer
c	40	41.4	1.5	278	22	AAH69999	Human cervical can
	41	41.4	1.5	773	22	AAH64909	Human secreted pro
	42	41.4	1.5	930	22	AAH33866	Human colon cancer
c	43	41.4	1.5	1664976	19	AAV21209	Methanococcus jann
	44	41.2	1.5	2106	22	AAF99980	Nucleotide sequenc
	45	41.2	1.5	2907	21	AAC98917	Human pancreatic c

ALIGNMENTS

RESULT 1
ABL12159
ID ABL12159 standard; cDNA; 3224 BP.
XX
AC ABL12159;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 30959.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB68056.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more

```
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 30959; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3224 BP; 844 A; 803 C; 808 G; 769 T; 0 other;


  Query Match             2.3%;  Score 61.8;  DB 23;  Length 3224;
  Best Local Similarity   46.4%;  Pred. No. 3.9e-06;
  Matches  241;  Conservative    0;  Mismatches  272;  Indels    6;  Gaps    1;

Qy     937 aaaatggaggatgtacaaccaattgatcgttttgtcgtggaggagtatttgctggatgtg 996
           ||  ||  || ||  ||  |||| || || ||| |  | |||||| || | |   || |
Db    1181 aatctgccgggtgcccactcaatagagcgctttttgattgaggagcatctacaccatatc 1240

Qy     997 ctcttctatttgaatggatgtcggaaggagtgtgcatcctacatggctaatcttcc---- 1052
           |       |  ||      || ||||| || ||  || |  ||  ||   ||||
Db    1241 attgaaacataccatcacgagcgaaaggattgcgccgcccagctgcttagctttccgtac 1300

Qy    1053 --tgttacatttcggtacgagtatcttatggcagagacactattttctcagatactgctg 1110
               |  | | |   |||||    || |  |||    | ||  |  || |  | |
Db    1301 aagcacaagatcccgctggagtactgcatcgtcgaggtggtcttcgccgagctctttcat 1360

Qy    1111 ctaccccagccaccattcaagactctttattatacactcgtgattatggatctttgtaag 1170
           | |||  ||| |  |    |    | |  ||      | |  | || || || || ||
Db    1361 atgcccacgccgcgctatctggacatctgctacggctccattttgatcgaactctgcaaa 1420

Qy    1171 gctcttccgggtgcctttcctgctgttgttgctggcgctgttcgtgcactatttgagaaa 1230
              | ||||   |  | ||    ||| | ||    ||          || |||  |
Db    1421 ctgcagccggccacgctgccccaggttctggcccaggccaccgagatcctctttatgcgc 1480

Qy    1231 atatccgacttagacatggaatccaggacgcgtcttatcctctggttttctcaccactta 1290
           ||   |    | ||| |  ||        ||  |  ||   ||||| ||  | ||  |
Db    1481 atcgactcgatgaacacgtcatgtttcgatcggttcgtcaattggttctcctatcatctg 1540

Qy    1291 tccaacttccaattcatctggccgtgggaagagtgggcttttgtgttggatcttcccaag 1350
             ||||||| | ||||  ||| | ||||| || ||||          ||   ||   |
Db    1541 agcaacttcaagttcacatggtcatgggacgaatgggacagctgtctgctcctcgacggc 1600

Qy    1351 tgggcccctaagcgtgtatttgttcaggagattttgcaaagagaagtacgcttgtcttac 1410
             |   ||     |    |||  | |||||| | ||||| |      |||||  | || ||
Db    1601 gagcatccgcgtcccaaattcatccaggaggtgttgcagaagtgcctacgcctttcgtat 1660

Qy    1411 tgggataaaattaagcagagcattgagaatgcgactgcc 1449
              |    |||| | | |  ||     |   |   ||||
Db    1661 catcagcgcattacggaaatgatgccaacgacttatgcc 1699


RESULT   2
ABL27678/c
ID   ABL27678 standard; DNA; 2902 BP.
XX
AC   ABL27678;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 34507.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 34507; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2902 BP; 727 A; 803 C; 769 G; 603 T; 0 other;


  Query Match             2.1%;  Score 56.8;  DB 23;  Length 2902;
  Best Local Similarity   46.6%;  Pred. No. 8.5e-05;
  Matches  221;  Conservative    0;  Mismatches  247;  Indels    6;  Gaps    1;

Qy     937 aaaatggaggatgtacaaccaattgatcgttttgtcgtggaggagtatttgctggatgtg 996
           ||  ||  || ||  ||  |||| || || ||| |  | |||||| || | |   || |
Db    2654 AATCTGCCGGGTGCCCACTCAATAGAGCGCTTTTTGATTGAGGAGCATCTACACCATATC 2595

Qy     997 ctcttctatttgaatggatgtcggaaggagtgtgcatcctacatggctaatcttcc---- 1052
           |       |  ||      || ||||| || ||  || |  ||  ||   ||||
Db    2594 ATTGAAACATACCATCACGAGCGAAAGGATTGCGCCGCCCAGCTGCTTAGCTTTCCGTAC 2535

Qy    1053 --tgttacatttcggtacgagtatcttatggcagagacactattttctcagatactgctg 1110
               |  | | |   |||||    || |  |||    | ||  |  || |  | |
Db    2534 AAGCACAAGATCCCGCTGGAGTACTGCATCGTCGAGGTGGTCTTCGCCGAGCTCTTTCAT 2475

Qy    1111 ctaccccagccaccattcaagactctttattatacactcgtgattatggatctttgtaag 1170
           | |||  ||| |  |    |    | |  ||      | |  | || || || || ||
Db    2474 ATGCCCACGCCGCGCTATCTGGACATCTGCTACGGCTCCATTTTGATCGAACTCTGCAAA 2415

Qy    1171 gctcttccgggtgcctttcctgctgttgttgctggcgctgttcgtgcactatttgagaaa 1230
              | ||||   |  | ||    ||| | ||    ||          || |||  |
Db    2414 CTGCAGCCGGCCACGCTGCCCCAGGTTCTGGCCCAGGCCACCGAGATCCTCTTTATGCGC 2355

Qy    1231 atatccgacttagacatggaatccaggacgcgtcttatcctctggttttctcaccactta 1290
           ||   |    | ||| |  ||        ||  |  ||   ||||| ||  | ||  |
Db    2354 ATCGACTCGATGAACACGTCATGTTTCGATCGGTTCGTCAATTGGTTCTCCTATCATCTG 2295

Qy    1291 tccaacttccaattcatctggccgtgggaagagtgggcttttgtgttggatcttcccaag 1350
             ||||||| | ||||  ||| | ||||| || ||||          ||   ||   |
Db    2294 AGCAACTTCAAGTTCACATGGTCATGGGACGAATGGGACAGCTGTCTGCTCCTCGACGGC 2235

Qy    1351 tgggcccctaagcgtgtatttgttcaggagattttgcaaagagaagtacgcttg 1404
             |   ||   .|    |||  | |||||| | ||||| |      |||| | |
```

```
Db    2234 GAGCATCCGCGTCCCAAATTCATCCAGGAGGTGTTGCAGAAGTGCCTACGGTAG 2181

RESULT   3
ABL13529
ID   ABL13529 standard; cDNA; 2244 BP.
XX
AC   ABL13529;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 35069.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69426.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 35069; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2244 BP; 509 A; 560 C; 626 G; 549 T; 0 other;

  Query Match             1.9%;  Score 50.8;  DB 23;  Length 2244;
  Best Local Similarity   49.3%;  Pred. No. 0.0032;
  Matches  133;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy    1060 tttcggtacgagtatcttatggcagagacactattttctcagatactgctgctaccccag 1119
           || | |  | | ||    || |  |||    | ||     |||| |||| ||| |||
Db    1054 ttgccggtctattactgcattgtggaggtgatcttgggcgagatgctgcggctgcccacc 1113

Qy    1120 ccaccattcaagactctttattatacactcgtgattatggatctttgtaaggctcttccg 1179
           |   | | ||  ||   ||   | | |  | |||| || || |||   |  |||
Db    1114 gccaattggagcaccattgcctacggatccattctggtggagctgtgcaagcgccaaccg 1173

Qy    1180 ggtgcctttcctgctgttgttgctggcgctgttcgtgcactatttgagaaaatatccgac 1239
           |     | ||   || || ||    ||  |   |  | | | | |   | |   | |
Db    1174 gacaagataccccaggtggtggcccaggccttggatataatctataatcgactgaactcc 1233

Qy    1240 ttagacatggaatccaggacgcgtcttatcctctggttttctcaccacttatccaacttc 1299
           |   | | |  | |     || ||  ||  |||| | || || ||| |   |||||||

Db    1234 atgagcgtagcttgcttcgaccgcctggtcaactgggtgtcgcatcacatcagcaacttc 1293

Qy    1300 caattcatctggccgtgggaagagtgggct 1329
              || | ||| | ||||    ||||||||
Db    1294 gggtttaactgccagtggagcaagtgggct 1323


RESULT   4
ABL13528
ID   ABL13528 standard; cDNA; 4420 BP.
XX
AC   ABL13528;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 35066.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69425.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 35066; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4420 BP; 1176 A; 907 C; 1033 G; 1304 T; 0 other;

  Query Match             1.9%;  Score 50.8;  DB 23;  Length 4420;
  Best Local Similarity   49.3%;  Pred. No. 0.0045;
  Matches  133;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy    1060 tttcggtacgagtatcttatggcagagacactattttctcagatactgctgctaccccag 1119
           || | |  | | ||    || |  |||    | ||     |||| |||| ||| |||
Db    2054 ttgccggtctattactgcattgtggaggtgatcttgggcgagatgctgcggctgcccacc 2113

Qy    1120 ccaccattcaagactctttattatacactcgtgattatggatctttgtaaggctcttccg 1179
           |   | | ||  ||   ||   | | |  | |||| || || |||   |  |||
Db    2114 gccaattggagcaccattgcctacggatccattctggtggagctgtgcaagcgccaaccg 2173

Qy    1180 ggtgcctttcctgctgttgttgctggcgctgttcgtgcactatttgagaaaatatccgac 1239
           |     | ||   || || ||    ||  |   |  | | | | |   | |   | |
```

```
Db   2174 gacaagataccccaggtggtggcccaggccttggatataatctataatcgactgaaactcc 2233

Qy   1240 ttagacatggaatccaggacgcgtcttatcctctggttttctcaccacttatccaacttc 1299
          |   | | |  | |      || ||  ||  |||| | || || ||| |   |||||||
Db   2234 atgagcgtagcttgcttcgaccgcctggtcaactgggtgtcgcatcacatcagcaacttc 2293

Qy   1300 caattcatctggccgtgggaagagtgggct 1329
             || | ||| | ||||    ||||||||
Db   2294 gggtttaactgccagtggagcaagtgggct 2323


RESULT   5
ABL34148/c
ID   ABL34148 standard; DNA; 6242 BP.
XX
AC   ABL34148;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2121.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 2121; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 6242 BP; 1699 A; 148 C; 1349 G; 3046 T; 0 other;


  Query Match             1.7%;  Score 45.4;  DB 24;  Length 6242;
  Best Local Similarity   52.4%;  Pred. No. 0.16;
  Matches  100;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy   2525 atgtgcatcctctctttctccaagccatatcttctgcacttcaattcccattacattaat 2584
          || |  |  ||||||||  |    ||| |  || | | ||  |   |  | | || |||
Db   1752 ATTTTTAATCTCTCTTTAACACCACCAAACTTTTTCCTCTCTATACCTTAAAATATAAAT 1693

Qy   2585 cttcctctttcaatctcaatcaaacctgtctcttttgtttttgttatgagattctgatt 2644
          || || | | | | |  ||| ||    |  ||||  | |     | | | | | |
Db   1692 TTTACTATCTAATTTTTCATCTAATAATTTCCTTTAATATACAAATTTAAAACTATTTAA 1633

Qy   2645 ctgacatcaagttattaggaaattgaaaagagtcaaaaaacaagagtttaaactttaaaa 2704
          | |||| | |   |       ||   |||  |   ||||   |  | | |||| |  ||||
Db   1632 CCGACAACTACCAAAACAATAACACAAATAATCAAAAATTTACAAATCTAAAATAAAAAA 1573

Qy   2705 aaaaaaaaaaa 2715
          ||||||||||||
Db   1572 AAAAAAAAAAA 1562


RESULT   6
AAV11855
ID   AAV11855 standard; cDNA; 1386 BP.
XX
AC   AAV11855;
XX
DT   11-SEP-1998  (first entry)
XX
DE   Homo sapiens Tub Interactor (hTI-1) gene.
XX
KW   serine protease; tub interactor; treatment; obesity; cachexia;
KW   anorexia nervosa; diabetes; cell cycle progression; apoptosis;
KW   neurodegenerative disease; Alzheimer's disease; drug screening;
KW   Parkinson's disease; Huntington's chorea; detection; diagnosis;
KW   amylotrophic lateral sclerosis; spinocerebellar degeneration; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..701
FT                   /*tag=  a
FT                   /product= hTI-1 protein
FT                   /note= "putative serine protease"
XX
PN   WO9812302-A1.
XX
PD   26-MAR-1998.
XX
PF   05-SEP-1997;   97WO-US15627.
XX
PR   21-JUL-1997;   97US-0897340.
PR   17-SEP-1996;   96US-0715032.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Errada PR,  Gimeno CJ;
XX
DR   WPI; 1998-217246/19.
DR   P-PSDB; AAW59129.
XX
PT   Tub interactor genes  - used to develop products for the treatment
PT   of obesity, cachexia, anorexia nervosa or related disorders e.g.
PT   diabetes
XX
PS   Claim 10; Fig 1; 120pp; English.
XX
CC   The sequence is that of the Tub Interactor gene hTI-1 which
CC   codes for a putative serine protease. TI genes function
CC   in biochemical pathways involved in weight control and
CC   related disorders.  The products can be used for treating
CC   weight disorders, e.g. obesity, cachexia or anorexia nervosa,
CC   or a related disorder such as diabetes.  The products can
CC   also be used to modulate cell cycle progression and apoptosis.
CC   They can be used for treating neurodegenerative diseases
CC   which are characterised by apoptosis, including Alzheimer's
CC   disease, Parkinson's disease, Huntington's chorea, amylotrophic
CC   lateral sclerosis or spinocerebellar degenerations. The
CC   products can also be used for detection, diagnosis and
CC   drug screening.
XX
```

```
SQ   Sequence 1386 BP; 318 A; 490 C; 321 G; 249 T; 8 other;


  Query Match             1.7%;  Score 45;  DB 19;  Length 1386;
  Best Local Similarity   55.1%;  Pred. No. 0.095;
  Matches   81;  Conservative    2;  Mismatches   64;  Indels    0;  Gaps    0;

Qy   2570 tcccattacattaatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgt 2629
        |||| ||||     |  ||| |   |  |  | | |||||  |   || ||||||  |
Db   1219 tccctgtacanattgccnccttgtggcacgttgacccaaccttaccagttggtttttcat 1278

Qy   2630 tatgagattctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaaga 2689
          | |  |   ||    |  | ||| ||  || | |     |: :|| | ||||||| || |
Db   1279 tttttgtccctttcccctagatccagaaataaagtttaagrgragsgccaaaaaaaaaaa 1338

Qy   2690 gtttaaactttaaaaaaaaaaaaaaa 2716
             |||    ||||||||||||||||
Db   1339 aaaaaaaaaaaaaaaaaaaaaaaaaa 1365


RESULT   7
AAS46678/c
ID   AAS46678 standard; DNA; 6552 BP.
XX
AC   AAS46678;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #400.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer  -
XX
PS   Claim 1; SEQ ID No 400; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes. Sequences with even numbered Seq ID numbers are the
CC   complementary sequence of the corresponding odd numbered sequence (e.g.
CC   ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC   is missing).
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6552 BP; 1584 A; 293 C; 1659 G; 3016 T; 0 other;


  Query Match             1.6%;  Score 44.4;  DB 22;  Length 6552;
  Best Local Similarity   56.0%;  Pred. No. 0.31;
  Matches   84;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy   2567 aattcccattacattaatcttcctctttcaatctcaatcaaacctgtctcttttgttttt 2626
          |||    || |  ||  | ||  | |||| || |   |      | | | |||| |||||
Db    621 AATAAAAATAATTTTTTTTTTTTTTTTCTATTTTCTTTCTTTTTTTTTTTTTTTTTTT 562

Qy   2627 tgttatgagattctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaaca 2686
          |  |     |   |  |||| |   ||| || | |  | ||  |||| | ||  |||| |
Db    561 TAATCAACAACCGTTTTTCTAATTCCAAATTTTAAATACATAAAAAAAAATCCGAAAAAA 502

Qy   2687 agagtttaaactttaaaaaaaaaaaaaaaa 2716
             |  | |||    ||||||||||||||||
Db    501 CCATATAAAAAAACAAAAAAAAAAAAAAA 472


RESULT   8
AAS46599
ID   AAS46599 standard; DNA; 6189 BP.
XX
AC   AAS46599;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #321.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer  -
```

XX
PS Claim 1; SEQ ID No 321; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6189 BP; 1750 A; 36 C; 1195 G; 3208 T; 0 other;

Query Match 1.6%; Score 44; DB 22; Length 6189;
Best Local Similarity 59.7%; Pred. No. 0.38;
Matches 74; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 2592 tttcaatctcaatcaaacctgtctcttttgtttttttgttatgagattctgattctgacat 2651
Db 2062 tttagagataaatgaaaaatgtagatttttttttattattaaaaaattttaaaattaaaat 2121

Qy 2652 caagttattaggaaattgaaaagagtcaaaaaacaagagtttaaactttaaaaaaaaaaa 2711
Db 2122 ttagatattaaaaaatttaaataataataaaatttttaaattaaagaaaaaaaaaaatt 2181

Qy 2712 aaaa 2715
Db 2182 taaa 2185

RESULT 9
ABL34592
ID ABL34592 standard; DNA; 6189 BP.
XX
AC ABL34592;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human metastasis associated gene SEQ ID NO: 145.
XX
KW Metastasis associated gene; cytostatic; gene therapy; cancer;
KW cytosine methylation; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200177376-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03970.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-010922/01.
XX
PT New nucleic acid derived from chemically treated metastasis genes,
PT useful for diagnosis of cancers by analysis of cytosine methylation,
PT also for treatment -
XX
PS Claim 1; SEQ ID NO 145; 23pp + Sequence Listing; English.
XX
CC The present invention provides a number of human metastasis associated
CC genes which are modified by cytosine methylation. The sequences can be
CC used in the diagnosis and treatment of cancer. The present sequence is
CC one of the genes of the invention.
XX
SQ Sequence 6189 BP; 1750 A; 36 C; 1195 G; 3208 T; 0 other;

Query Match 1.6%; Score 44; DB 24; Length 6189;
Best Local Similarity 59.7%; Pred. No. 0.38;
Matches 74; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 2592 tttcaatctcaatcaaacctgtctcttttgtttttttgttatgagattctgattctgacat 2651
Db 2062 tttagagataaatgaaaaatgtagatttttttttattattaaaaaattttaaaattaaaat 2121

Qy 2652 caagttattaggaaattgaaaagagtcaaaaaacaagagtttaaactttaaaaaaaaaaa 2711
Db 2122 ttagatattaaaaaatttaaataataataaaatttttaaattaaagaaaaaaaaaaatt 2181

Qy 2712 aaaa 2715
Db 2182 taaa 2185

RESULT 10
AAL15089/c
ID AAL15089 standard; cDNA; 669 BP.
XX
AC AAL15089;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 7546.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX

```
PS   Claim 1; Page 1358; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 669 BP; 260 A; 81 C; 74 G; 166 T; 88 other;


  Query Match             1.6%;  Score 43.6;  DB 22;  Length 669;
  Best Local Similarity   47.3%;  Pred. No. 0.16;
  Matches   88;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy    2530 catcctctctttctccaagccatatcttctgcacttcaattcccattacattaatcttcc 2589
           | || | | ||| |       | | || |  |    ||||    ||   ||    ||
Db     270 CNTCTTTTTTTTTTTNNTNTGNTNTNTTATNAAAAANAATTTNTTTTTTTTTTTNTTTNN 211

Qy    2590 tctttcaatctcaatcaaacctgtctcttttgttttttgttatgagattctgattctgac 2649
             |     | |   |       | | | |||| |||||| ||      |   |  | |
Db     210 NNTNNGNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNCTTTTNGGGCCTNTCNA 151

Qy    2650 atcaagttattaggaaattgaaaagagtcaaaaaacaagagtttaaactttaaaaaaaaa 2709
           || ||   |    |||   |||| |   |||||| || |    |||    |||||||||
Db     150 ATTAAAANAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 91

Qy    2710 aaaaaa 2715
           ||||||
Db      90 AAAAAA 85


RESULT  11
AAQ25449
ID   AAQ25449 standard; DNA; 2872 BP.
XX
AC   AAQ25449;
XX
DT   07-DEC-1992  (first entry)
XX
DE   Cholesterol 7-alpha-hydrolase gene.
XX
KW   C7H; human; 7-hydroxycholesterol; recombinant; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             40..504
FT                   /*tag=  a
XX
PN   JP04144680-A.
XX
PD   19-MAY-1992.
XX
PF   05-SEP-1990;   90JP-0268587.
XX
PR   05-SEP-1990;   90JP-0268587.
XX
PA   (TOXN ) TOYO JOZO KK.
XX
DR   WPI; 1992-214120/26.
DR   P-PSDB; AAR24482.
XX
PT   Cholesterol 7-alpha-hydroxylase enzyme - and DNA sequence etc., for
PT   efficient prodn. of enzyme
XX
PS   Claim 1; Fig 1; 15pp; Japanese.
XX
CC   The DNA sequence encodes human cholesterol 7-alpha-hydrolase which
CC   catalyses the formation of 7-alpha-hydroxycholesterol, NADP and
CC   water from cholesterol, reduced NADP and molecular state oxygen.
CC   The enzyme may be produced by recombinant techniques by culturing
CC   cells transformed with a plasmid contg. the DNA sequence and
CC   recovering the enzyme from the culture.  C7H may be produced
CC   efficiently by this method.
XX
SQ   Sequence 2872 BP; 898 A; 532 C; 561 G; 881 T; 0 other;


  Query Match             1.6%;  Score 43.4;  DB 13;  Length 2872;
  Best Local Similarity   53.9%;  Pred. No. 0.38;
  Matches   89;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    2552 tatcttctgcacttcaattcccattacattaatcttcctctttcaatctcaatcaaacct 2611
           ||||  ||  |  ||||||   ||| | |    |   | | |  |  | |     ||||
Db    2704 tatcagctctatgtcaattgatatttctgtgtagtatttatatttaaattatatttacct 2763

Qy    2612 gtctcttttgtttttttgttatgagattctgattctgacatcaagttattaggaaattgaa 2671
             | |  ||  |||     ||| |       ||||| ||||    | |||  |||||
Db    2764 ttttgcttattttacaaatattaagaaaatattctaacatttgataattttgaaatgatt 2823

Qy    2672 aagagtcaaaaaacaagagtttaaactttaaaaaaaaaaaaaaa 2716
           |   |  | ||| || ||| | ||  | ||||||||||||||||
Db    2824 catctttcagaaataaaagtatgaatctaaaaaaaaaaaaaaaa 2868


RESULT  12
AAD03657
ID   AAD03657 standard; DNA; 1456 BP.
XX
AC   AAD03657;
XX
DT   19-JUN-2001  (first entry)
XX
DE   Human serine protease, prost 07 DNA.
XX
KW   Human; serine protease; PROST 07; gene therapy; vaccine; cytostatic;
KW   tumour metastasis; ribozyme therapy; antisense therapy; cancer; prostate;
KW   ovary; endometrium; benign prostatic hyperplasia; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..705
FT                   /*tag=  a
FT                   /product= "Human serine protease, PROST 07"
FT   sig_peptide     1..30
FT                   /*tag=  b
FT                   /note= "Encodes pro-sequence and part of a signal
FT                   sequence of human serine protease, PROST 07"
FT   mat_peptide     31..703
FT                   /*tag=  c
FT                   /label= "Mature human serine protease, PROST 07"
XX
PN   WO200125446-A1.
XX
PD   12-APR-2001.
XX
PF   05-OCT-2000; 2000WO-US27431.
XX
PR   07-OCT-1999;   99US-0158588.
PR   02-OCT-2000; 2000US-0678940.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Bringmann PW,  Brink J,  Harkins R,  Light DR,  Lin RJ,  Parkes D;
PI   Parry G,  Schneider DW,  Steinbrecher R,  Van Heuit PT,  Xuan J;
XX
DR   WPI; 2001-266314/27.
DR   P-PSDB; AAE00397.
XX
PT   Novel serine protease polypeptide and polynucleotides, designated PROST
```

PT 07 useful in research, diagnosis, treatment of cell proliferation
PT diseases, such as prostate and ovarian cancer -
XX
PS Claim 8; Fig 1; 88pp; English.
XX
CC The present cDNA sequence encodes human serine protease, designated as
CC PROST 07. Prost 07 DNA is found in cDNA clones obtained from human
CC prostate tissues. PROST 07 nucleotides are used as DNA probes and as
CC targets for antisense and ribozyme therapy. PROST 07 shows amino acid
CC sequence homology to human kallikrein family of serine protease. PROST 07
CC is useful in the diagnosis and treatment of diseases of cell
CC proliferation, such as prostate cancer, benign prostatic hyperplasia,
CC ovarian cancer and endometrial cancer. PROST 07 is also useful for
CC generating antibodies which are useful in detecting the presence of
CC prostate cancer and tumour metastasis. It is also used to provide
CC vaccines for generating humoral and cell-mediated immunity for use in
CC anti-cancer therapy. PROST 07 antibodies are also useful in imaging
CC methods. PROST 07 is used as an excellent target for therapeutic methods
CC such as targetted antibody therapy, immunotherapy and gene therapy.
XX
SQ Sequence 1456 BP; 375 A; 504 C; 326 G; 251 T; 0 other;

Query Match 1.6%; Score 43; DB 22; Length 1456;
Best Local Similarity 55.8%; Pred. No. 0.34;
Matches 82; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 2570 tcccattacattaatcttcctctttcaatctcaatcaaacctgtctcttttgtttttgt 2629
Db 1247 tccctgtacacagtgccccttgtggcacgttgacccaaccttaccagttggtttttcat 1306

Qy 2630 tatgagattctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaaga 2689
Db 1307 tttttgtccctttcccctagatccagaaataaagtctaagagaagcgcaaaaaaaaaaaa 1366

Qy 2690 gtttaaactttaaaaaaaaaaaaaaa 2716
Db 1367 aaaaaaaaaaaaaaaaaaaaaaaaaa 1393

RESULT 13
AAS07790/c
ID AAS07790 standard; DNA; 256 BP.
XX
AC AAS07790;
XX
DT 23-OCT-2001 (first entry)
XX
DE Cervical cancer pre-malignant condition DNA marker #89.
XX
KW Cervical cancer; pre-malignant condition marker; tumourigenesis; CIN; ds;
KW cervical intraepithelial neoplasia; squamous intraepithelial lesions;
KW polymerase chain reaction; PCR; enzyme linked immunosorbent assay; ELISA;
KW anticancer therapy; carcinogen; antisense inhibition.
XX
OS Homo sapiens.
XX
PN WO200142792-A2.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33311.
XX
PR 08-DEC-1999; 99US-0169811.
PR 21-DEC-1999; 99US-0171330.
PR 14-MAR-2000; 2000US-0189113.
PR 31-MAR-2000; 2000US-0193943.
PR 12-MAY-2000; 2000US-0203772.
PR 09-JUN-2000; 2000US-0210820.
PR 21-JUL-2000; 2000US-0220113.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI; 2001-367889/38.
XX
PT Cervical cancer protein markers useful for the diagnosis, prevention
PT and treatment of cervical cancers, especially cervical intraepithelial
PT neoplasia or squamous intraepithelial lesions -
XX
PS 20; Page 413; 436pp; English.
XX
CC The sequence represents a cervical cancer pre-malignant condition marker.
CC These markers encode proteins that are over expressed during
CC tumourigenesis. The proteins and their corresponding nucleic acid
CC sequences can therefore be used for the diagnosis, prevention and
CC treatment of cervical cancers, particularly cervical intraepithelial
CC neoplasia (CIN) and squamous intraepithelial lesions (SIL). The sequences
CC may be used as markers in diagnostic assays to detect cancerous
CC conditions (e.g. by polymerase chain reaction (PCR) or enzyme linked
CC immunosorbent assay (ELISA)), to monitor the efficacy of anticancer
CC therapies and to identify anticancer or carcinogenic compounds. The level
CC of expression of a marker in a patient sample is compared with the normal
CC level of expression of the marker in a control non-cervical cancer
CC sample, whereby a significant difference indicates that the patient is
CC afflicted with cervical cancer or a pre-malignant condition. The
CC expression of these proteins may be inhibited by antisense inhibition for
CC the treatment of cancers. They may be used in this way for the treatment
CC of cervical intraepithelial neoplasia or squamous intraepithelial
CC lesions.
XX
SQ Sequence 256 BP; 69 A; 18 C; 31 G; 122 T; 16 other;

Query Match 1.6%; Score 42.8; DB 22; Length 256;
Best Local Similarity 51.0%; Pred. No. 0.16;
Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 2560 gcacttcaattcccattacattaatcttcctctttcaatctcaatcaaacctgtctcttt 2619
Db 248 GGNCCCCNATTTTTTTTTTTTTTTTTTTCCCCCCCTTTTTTTTTTCCCNTTTTTTTTT 189

Qy 2620 tgtttttgttatgagattctgattctgacatcaagttattaggaaattgaaaagagtca 2679
Db 188 TTTTTTTTTTTTNNNGGTTTTNNNAAANTTTTNAANTTTTTTNNGNAAAAAAAAAAAAAA 129

Qy 2680 aaaaacaagagtttaaactttaaaaaaaaaaaaaaa 2716
Db 128 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 92

RESULT 14
AAH71471
ID AAH71471 standard; cDNA; 612 BP.
XX
AC AAH71471;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 2745.
XX
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200142467-A2.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33312.
XX
PR 08-DEC-1999; 99US-0169681.
PR 21-DEC-1999; 99US-0171350.
PR 14-MAR-2000; 2000US-0189315.
PR 12-MAY-2000; 2000US-0203791.

PR 09-JUN-2000; 2000US-0210600.
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1; Page 564; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 612 BP; 380 A; 50 C; 2 G; 178 T; 2 other;

Query Match 1.6%; Score 42.8; DB 22; Length 612;
Best Local Similarity 56.3%; Pred. No. 0.25;
Matches 80; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 2575 ttacattaatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgttatga 2634
Db 22 tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 81

Qy 2635 gattctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaagagttta 2694
Db 82 ttttttaattttttttttaattaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 141

Qy 2695 aactttaaaaaaaaaaaaaaaa 2716
Db 142 aaaaataaaaaaaaaaaaaaaa 163

RESULT 15
AAH71471/c
ID AAH71471 standard; cDNA; 612 BP.
XX
AC AAH71471;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 2745.
XX
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200142467-A2.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33312.
XX
PR 08-DEC-1999; 99US-0169681.
PR 21-DEC-1999; 99US-0171350.
PR 14-MAR-2000; 2000US-0189315.
PR 12-MAY-2000; 2000US-0203791.
PR 09-JUN-2000; 2000US-0210600.
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1; Page 564; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 612 BP; 380 A; 50 C; 2 G; 178 T; 2 other;

Query Match 1.6%; Score 42.8; DB 22; Length 612;
Best Local Similarity 56.3%; Pred. No. 0.25;
Matches 80; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 2575 ttacattaatcttcctctttcaatctcaatcaaacctgtctcttttgttttttgttatga 2634
Db 182 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTT 123

Qy 2635 gattctgattctgacatcaagttattaggaaattgaaaagagtcaaaaaacaagagttta 2694
Db 122 TTTTTTTTTTTTTTTTTTTAATTAAAAAAAAAATTAAAAAAAAAAAAAAAAAAAAAAAAA 63

Qy 2695 aactttaaaaaaaaaaaaaaaa 2716
Db 62 AAAAAAAAAAAAAAAAAAAAAA 41

Search completed: August 20, 2002, 13:58:51
Job time: 8270 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 20, 2002, 15:25:57 ; Search time 87.13 Seconds
(without alignments)
1683.688 Million cell updates/sec

Title: US-09-882-986-2
Perfect score: 4374
Sequence: 1 MSNWKTLLLRIGEKGPEYGT.....EDVHPLFLQAISSALQFPLH 848

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4374	100.0	848	10	Q9LKN6	Q9lkn6 arabidopsis
2	3784	86.5	749	10	Q9SIU2	Q9siu2 arabidopsis
3	2753.5	63.0	910	10	Q9AWB3	Q9awb3 oryza sativ
4	818.5	18.7	800	5	Q9U980	Q9u980 drosophila
5	818.5	18.7	800	5	Q9W4I2	Q9w4i2 drosophila
6	636.5	14.6	754	5	O01763	O01763 caenorhabdi
7	550.5	12.6	747	5	Q9VDA4	Q9vda4 drosophila
8	501.5	11.5	829	3	Q9HE91	Q9he91 neurospora
9	475.5	10.9	780	3	O14253	O14253 schizosacch
10	374.5	8.6	886	3	Q9Y717	Q9y717 candida alb
11	163	3.7	1413	5	Q9XWR0	Q9xwr0 caenorhabdi
12	145.5	3.3	1833	5	Q9VM67	Q9vm67 drosophila
13	145	3.3	1023	16	Q92B64	Q92b64 listeria in
14	142.5	3.3	802	4	Q96JN1	Q96jn1 homo sapien
15	142	3.2	2473	11	Q9QZ84	Q9qz84 mus musculu
16	140	3.2	906	4	Q96ST0	Q96st0 homo sapien
17	139.5	3.2	1937	13	Q9IBD4	Q9ibd4 gallus gall
18	139.5	3.2	3259	4	Q14789	Q14789 homo sapien
19	139	3.2	792	10	Q9ARQ9	Q9arq9 oryza sativ
20	139	3.2	1388	13	Q91785	Q91785 xenopus lae
21	138.5	3.2	3868	5	Q9W090	Q9w090 drosophila
22	138	3.2	1295	5	Q22257	Q22257 caenorhabdi
23	138	3.2	1930	13	Q9DGD5	Q9dgd5 pennahia ar
24	137	3.1	1156	16	O66878	O66878 aquifex aeo
25	136	3.1	1133	5	Q21022	Q21022 caenorhabdi
26	136	3.1	3419	11	O55147	O55147 rattus norv
27	135	3.1	4151	5	O96936	O96936 drosophila
28	135	3.1	5201	5	Q9U479	Q9u479 drosophila
29	135	3.1	5380	11	Q91ZU7	Q91zu7 mus musculu
30	135	3.1	5385	5	Q9V6V3	Q9v6v3 drosophila
31	135	3.1	5496	5	Q9V6V2	Q9v6v2 drosophila
32	135	3.1	7389	11	Q91ZU6	Q91zu6 mus musculu
33	135	3.1	8805	5	Q9V6V4	Q9v6v4 drosophila
34	134.5	3.1	1084	4	Q9NVI1	Q9nvi1 homo sapien
35	134.5	3.1	1362	4	Q9ULF8	Q9ulf8 homo sapien
36	134	3.1	1314	3	Q9C499	Q9c499 neurospora
37	133.5	3.1	775	4	Q96CN9	Q96cn9 homo sapien
38	133.5	3.1	2331	5	Q22847	Q22847 caenorhabdi
39	133.5	3.1	3263	5	Q9I7U3	Q9i7u3 drosophila
40	132.5	3.0	1171	3	Q9P3E2	Q9p3e2 neurospora
41	132.5	3.0	1627	5	Q962Q0	Q962q0 giardia lam
42	132	3.0	4063	5	O02425	O02425 caenorhabdi
43	131.5	3.0	809	11	Q9D4H2	Q9d4h2 mus musculu
44	131.5	3.0	1313	10	Q9XIP6	Q9xip6 arabidopsis
45	131.5	3.0	1864	10	Q9FYL7	Q9fyl7 arabidopsis

ALIGNMENTS

RESULT 1
Q9LKN6
ID Q9LKN6 PRELIMINARY; PRT; 848 AA.
AC Q9LKN6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NUCLEAR CAP-BINDING PROTEIN CBP80 (MRNA CAP BINDING PROTEIN).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Kmieciak M., Jarmolowski A.;
RT "A nuclear cap-binding protein complex from Arabidopsis thaliana.
RT Molecular characterization of Arabidopsis thaliana CBP80.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21417142; PubMed=11525733;
RA Hugouvieux V., Kwak J.M., Schroeder J.I.;
RT "An mRNA Cap Binding Protein, ABH1, Modulates Early Abscisic Acid
RT Signal Transduction in Arabidopsis.";
RL Cell 106:477-487(2001).
DR EMBL; AF268377; AAF76167.1; -.
DR EMBL; AF272891; AAK91588.1; -.
DR InterPro; IPR002106; AA_tRNA_ligase_II.
DR InterPro; IPR003890; EIF4G_cent.
DR SMART; SM00543; MIF4G; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
SQ SEQUENCE 848 AA; 96547 MW; 1E583391F4BAE1A8 CRC64;

Query Match 100.0%; Score 4374; DB 10; Length 848;
Best Local Similarity 100.0%; Pred. No. 9.9e-293;

```
  Matches  848;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MSNWKTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSNWKTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHK 60

Qy      61 IPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 IPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQP 120

Qy     121 ASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERV 180

Qy     181 LVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRP 240

Qy     241 HLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPAN 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 HLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPAN 300

Qy     301 KMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLP 360

Qy     361 QPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHLSN 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 QPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHLSN 420

Qy     421 FQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLPPKA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 FQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLPPKA 480

Qy     481 GPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQ 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQ 540

Qy     541 TLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMM 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMM 600

Qy     601 GYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLV 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLV 660

Qy     661 AEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKE 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKE 720

Qy     721 ALLNRALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSEN 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ALLNRALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSEN 780

Qy     781 GNPKKSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAIS 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GNPKKSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAIS 840

Qy     841 SALQFPLH 848
           ||||||||
Db     841 SALQFPLH 848


RESULT   2
Q9SIU2
ID   Q9SIU2      PRELIMINARY;      PRT;   749 AA.
AC   Q9SIU2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE CAP-BINDING PROTEIN.
GN   AT2G13540.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007063; AAD22677.1; -.
DR   InterPro; IPR002106; AA_tRNA_ligase_II.
DR   InterPro; IPR003890; EIF4G_cent.
DR   Pfam; PF02854; MIF4G; 1.
DR   SMART; SM00543; MIF4G; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
SQ   SEQUENCE   749 AA;  85610 MW;  8AF328F08CD33CB0 CRC64;


  Query Match             86.5%;  Score 3784;  DB 10;  Length 749;
  Best Local Similarity   99.1%;  Pred. No. 3.6e-252;
  Matches  735;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 MSNWKTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSNWKTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHK 60

Qy      61 IPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 IPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQP 120

Qy     121 ASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERV 180

Qy     181 LVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRP 240

Qy     241 HLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPAN 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 HLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPAN 300

Qy     301 KMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLP 360

Qy     361 QPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHLSN 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 QPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHLSN 420

Qy     421 FQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLPPKA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 FQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLPPKA 480

Qy     481 GPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQ 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQ 540

Qy     541 TLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMM 600
```

```
Db    541 TLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMM 600

Qy    601 GYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLV 660
Db    601 GYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLV 660

Qy    661 AEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKE 720
Db    661 AEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKE 720

Qy    721 ALLNRALSETEVLLLLLFQSFL 742
Db    721 ALLNRALSETEVLYISLMQHMI 742


RESULT   3
Q9AWB3
ID   Q9AWB3      PRELIMINARY;      PRT;   910 AA.
AC   Q9AWB3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NUCLEAR CAP-BINDING PROTEIN CBP80.
OS   Oryza sativa subsp. japonica.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Derba M., Kmieciak M., Jarmolowski A.;
RT   "Molecular characterization of Oryza sativa CBP80.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY017415; AAG54079.1; -.
DR   InterPro; IPR002106; AA_tRNA_ligase_II.
DR   InterPro; IPR003890; EIF4G_cent.
DR   InterPro; IPR002052; N6_Mtase.
DR   Pfam; PF02854; MIF4G; 1.
DR   SMART; SM00543; MIF4G; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
SQ   SEQUENCE   910 AA;  103823 MW;  23D62F81E8596F56 CRC64;


  Query Match           63.0%;  Score 2753.5;  DB 10;  Length 910;
  Best Local Similarity   61.3%;  Pred. No. 4.7e-181;
  Matches  524;  Conservative  136;  Mismatches  176;  Indels   19;  Gaps    5;

Qy      2 SNWKTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHKI 61
Db      3 AGWRTLLLRIGDRCPEYGGSVDHKEHIETCYGVLCREYEHSKDAMFEFLLQCADQLPHKI 62

Qy     62 PLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPA 121
Db     63 PFFGVLIGLINLENEDFSKGIVDTTHANLQDALHNENRDRIRILLRFLCGLMCSKVVLPN 122

Qy    122 SLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVL 181
Db    123 SIIETFEALLSSAATILDEETGNPSWQPRADFYVYCILASLPWGGSELFEQVPDEFERVL 182

Qy    182 VGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPH 241
Db    183 VGIQSYISIRRHFDDIAFSVFETDE-GNSPNKKDFIEDLWERIQVLSRNGWKVKSVPKPH 241

Qy    242 LSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANK 301
Db    242 LSFEAQLVAGVSHRFSPISCPHLLSRNAS-SEIVKGQEKHEADLKYPQRLRRLHIFPTNK 300

Qy    302 MEDVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLPQ 361
Db    301 AENMQPVDRFVVEECILDVLLFFNGCRKECAFYLVSLPVPFRYEYLMAETIFSQLLLLPN 360

Qy    362 PPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHLSNF 421
Db    361 PPFRPIYYTLVIIDLCKALPGAFPSVVVGAVHALFDRISNMDMECRTRLILWFSHHLSNF 420

Qy    422 QFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLPPKAG 481
Db    421 QFIWPWQEWAYVKDLPKWAPQRVFVQEVLEREIRLSYFDKIKQSIEDAVELEELLPPKAG 480

Qy    482 PNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQT 541
Db    481 PNFRYHSDEGKESTDGHRLSKELVAMVRGRKTQGDIISWVDEKIIPVNGAKFALDVVSQT 540

Qy    542 LLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMG 601
Db    541 LLDIGSKSFTHLITVLERYGQIISKLCPNEEMQLLLMDEVSAYWKNSTQMIAIAIDRMMG 600

Qy    602 YRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVA 661
Db    601 YRLLSNLAIVKWVFSPANVDQFHVSDRPWEILRNAVSKTYNRIFDLRKEIQTLRKGLQAA 660

Qy    662 EKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEA 721
Db    661 KEASEKAARELEEAKSIIEIVDGQPVPSENPGRLRRLQARADKAKEGEVTTEESLEAKEA 720

Qy    722 LLNRALSETEVLLLLLFQSFLGVLKERLP------DPTKVRSVQDLKSIGAEDDKPSAMD 775
Db    721 LLARGLEESKELLRLLFKSFVEVLTERLPPISADGDVPNLRAGDPNVNSSARDPEATTME 780

Qy    776 VDSENG----------NPKKSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEV 825
Db    781 IDNENGGDNDSQLNGQNKKISHNVGELEQWCLCTLGYLKSFSRQYATEIWSHIAMLDQEI 840

Qy    826 FSGEDVHPLFLQAIS 840
Db    841 FVG-NIHPLIRKAAS 854


RESULT   4
Q9U980
ID   Q9U980      PRELIMINARY;      PRT;   800 AA.
AC   Q9U980;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CAP BINDING PROTEIN 80.
GN   CBP80 OR EG:84H4.3 OR CG7035.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON R; TISSUE=IMAGINAL DISC;
RA   Lewis J.D.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ238970; CAB53186.1; -.
DR   FlyBase; FBgn0022942; Cbp80.
DR   InterPro; IPR003890; EIF4G_cent.
DR   Pfam; PF02854; MIF4G; 1.
DR   SMART; SM00543; MIF4G; 1.
SQ   SEQUENCE   800 AA;  93070 MW;  7C9E7948ABD876B6 CRC64;


  Query Match           18.7%;  Score 818.5;  DB 5;  Length 800;
  Best Local Similarity   25.6%;  Pred. No. 7.7e-48;
  Matches  218;  Conservative  175;  Mismatches  342;  Indels  115;  Gaps   19;

Qy      5 KTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHKIPLY 64
Db     35 ESLILRVGE-----ATTSSVESNLEGLVSVLEADLGTFRLKILRILSDCAVAMPEKCTVY 89

Qy     65 GTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLI 124
```

```
Db      90 TTLVGLLNAKNYKFGGEFVDHMVKTFKESLKMCRWDAARYSLRFLADLVNCHVISATSLL 149

Qy     125 VVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGI 184

Db     150 QLLDTMID-----VSNEDTVP--QVRRDWFVFAVLSTLPWVGRDLYEKKESALESLLLRI 202

Qy     185 QAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSF 244

Db     203 EVYLNKRSKKHHNALRVWSS---DAPHPQEEYLDCLWAQIRKLRQDNWAEKHIPRPYLVF 259

Qy     245 EAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMED 304

Db     260 DSILCEALQHNLPTI------VPPPHHDN-----------FEYPMPWVVYRMFDYTDCPD 302

Qy     305 ------VQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFR--YEYLMAETLFSQI 356

Db     303 GPNLPGAHSIERFLIEEHLHHIIETYHHERKDCAAQLLSFPYKHKIPLEYCIVEVVFAEL 362

Qy     357 LLLPQPPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSH 416

Db     363 FHMPTPRYLDICYGSILIELCKLQPATLPQVLAQATEILFMRIDSMNTSCFDRFVNWFSY 422

Qy     417 HLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELL 476

Db     423 HLSNFKFTWSWDEWDSCLLLDGEHPRPKFIQEVLQKCLRLSYHQRITEMM--PTTYAKLI 480

Qy     477 PPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFE---- 532

Db     481 PLTPVPNYKYANEEA-ANLPGTTVAHQLVVAIRQKCTPEEVVNILKD--IPNSGYSGEEM 537

Qy     533 -------VTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYW 585

Db     538 SDGSFNALKIDVFVQTLLNLGSKSFSHSFAAISKFHSVFRALAETEEAQICILHNIFELW 597

Qy     586 KNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRIS 645

Db     598 SSHQQMMVVLIDKLLKLQIVDCSAVATWIFSKEMTGEF-TKLYLWEILHLTIKKMNKHVI 656

Qy     646 DLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKT 705

Db     657 KLNTELSEAKEKLAKADSSSSDS--------------------EDDSSHKRKKPIT--- 692

Qy     706 GEAELSLRESLEAKEALLNRALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIG 765

Db     693 -HADKPSEEVVERMEEKLEAANVNQKRLFLIVFQRFIMILSEHL-----LRSDTD----- 741

Qy     766 AEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEV 825

Db     742 -------GRDPDTD--------------WYRWTIGRLQQVFLMHHEQVQKYSSTLETLL 779

Qy     826 FSGE-DVHPL 834

Db     780 FTSDLDTHIL 789


RESULT   5
Q9W4I2
ID   Q9W4I2      PRELIMINARY;      PRT;   800 AA.
AC   Q9W4I2; O77276;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CBP80 PROTEIN (EG:84H4.3 PROTEIN) (LD31211P).
GN   CBP80 OR EG:84H4.3 OR CG7035.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE OF 1-348 FROM N.A.
RA   Ferraz C., Vidal S., Brun C., Bucheton A., Demaille J.G.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-348 FROM N.A.
RA   Benos P.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, CN BW SP;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003431; AAF45970.1; -.
DR   EMBL; AL031766; CAA21136.1; -.
DR   EMBL; AY051796; AAK93220.1; -.
DR   FlyBase; FBgn0022942; Cbp80.
DR   InterPro; IPR003890; EIF4G_cent.
DR   Pfam; PF02854; MIF4G; 1.
DR   SMART; SM00543; MIF4G; 1.
SQ   SEQUENCE   800 AA;  93226 MW;  4D90C157E9A33B9B CRC64;


  Query Match             18.7%;  Score 818.5;  DB 5;  Length 800;
  Best Local Similarity   25.5%;  Pred. No. 7.7e-48;
  Matches  217;  Conservative  178;  Mismatches  340;  Indels  115;  Gaps   19;

Qy       5 KTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHKIPLY 64
```

```
Db     35 ESLILRVGER-----STSSVESNLEGLVSVLEADLGTFRLKILRILSDCAVRMPEKCTVY 89

Qy     65 GTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLI 124
          ||:|||| :|  |  : |: :  |: :|     :: |  |||:  |:   ||   ||:
Db     90 TTLVGLLNAKNYKFGGEFVDHMVKTFKESLKMCRWDAARYSLRFLADLVNCHVISATSLL 149

Qy    125 VVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGI 184
          : :|::      |  |   |  | : |::|  :||:||| | :| |:    :| :|: |
Db    150 QLLDTMID-----VSNEDTVP--QVRRDWFVFAVLSTLPWVGRDLYEKKESALESLLLRI 202

Qy    185 QAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSF 244
          : ||: |     : |  : :   ::   :::::: |  :|: |  : |  : :|||:| |
Db    203 EVYLNKRSKKHHNALRVWSS---DAPHPQEEYLDCLWAQIRKLRQDNWAEKHIPRPYLVF 259

Qy    245 EAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMED 304
          :: |     | |  |          ||  |              ||      :|      |
Db    260 DSILCEALQHNLPTI------VPPPHHDN-----------FEYPMPWVVYRMFDYTDCPD 302

Qy    305 ------VQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFR--YEYLMAETLFSQI 356
                |:||::||:|  ::   :  ||:||: : : |   :   || : | :|:::
Db    303 GPNLPGAHSIERFLIEEHLHHIIETYHHERKDCAAQLLSFPYKHKIPLEYCIVEVVFAEL 362

Qy    357 LLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSH 416
            :| | :  : |  ::::|||  |   | |:| |   || :|  ::      | : |||:
Db    363 FHMPTPRYLDICYGSILIELCKLQPATLPQVLAQATEILFMRIDSMNTSCFDRFVNWFSY 422

Qy    417 HLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELL 476
          |||||:| | |:||   | |    |:  |:||:||: :|||| :| : :   |   :|:
Db    423 HLSNFKFTWSWDEWDSCLLLDGEHPRPKFIQEVLQKCLRLSYHQRITEMM--PTTYAKLI 480

Qy    477 PPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFE---- 532
          |    ||: |:'||         :: :|   :::| |  :::  :::   |  |:
Db    481 PLTPVPNYKYANEEA-ANLPGTTVAHQLVVAIRQKCTPEEVVNILKD--IPNSGYSGEEM 537

Qy    533 -------VTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYW 585
                : : : |||||::|||||:|    : ::  ||  |    : |: :|  :   |
Db    538 SDGSFNALKIDVFVQTLLNLGSKSFSHSFAAISKFHSVFRALAETEEAQICILHNIFELW 597

Qy    586 KNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRIS 645
          :: ||  | ||:::  ::|   |:  |:|| |   :|      ||||   : |    :
Db    598 SSHQQMMVVLIDKLLKLQIVDCSAVATWIFSKEMTGEF-TKLYLWEILHLTIKKMNKHVI 656

Qy    646 DLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKT 705
          |  ::|   : :  |: :|:::                  |: :  || |
Db    657 KLNTELSEAKEKLAKADSSSSDS--------------------EDDSSHKRKKPIT--- 692

Qy    706 GEAELSLRESLEAKEALLNRALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIG 765
            |:    | :|  |  |  |    : | |::|| |: :| | |      :||  |
Db    693 -HADKPSEEVVERMEEKLEAANVNQKRLFLIVFQRFIMILSEHL-----LRSDTD----- 741

Qy    766 AEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEV 825
                | |::               |   |:| |      :  ::  :   |:| :
Db    742 -------GRDPDTD--------------WYRWTIGRLQQVFLMHHEQVQKYSSTLETLL 779

Qy    826 FSGE-DVHPL 834
          |: : | | |
Db    780 FTSDLDTHIL 789


RESULT   6
O01763
ID   O01763      PRELIMINARY;      PRT;   754 AA.
AC   O01763;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 87.7 KDA PROTEIN.
GN   F37E3.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Wamsley P., Bradshaw H.;
RT   "The sequence of C. elegans cosmid F37E3.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF003132; AAB54133.2; -.
DR   InterPro; IPR000179; Cyt_b_b6.
DR   InterPro; IPR003890; EIF4G_cent.
DR   Pfam; PF02854; MIF4G; 1.
DR   SMART; SM00543; MIF4G; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   754 AA;  87715 MW;  FFE4A943CBE23876 CRC64;


  Query Match             14.6%;  Score 636.5;  DB 5;  Length 754;
  Best Local Similarity   24.4%;  Pred. No. 2.4e-35;
  Matches  182;  Conservative  163;  Mismatches  322;  Indels   79;  Gaps   20;

Qy     11 IGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHKIPLYGTLIGL 70
          ||: |   |:| :   :::     :  ::|:    ::  :  ||  ||::: :| ||:||
Db     35 IGKVGENTGSSIEC--NLDKLTAFLHDDLEKYRASIIDIIAGCAIYLPNRVTVYTTLVGL 92

Qy     71 LNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLIVVFETL 130
          || :| :|   :|| : :  |  |     : |  |:  |  | |:   |:    |:
Db     93 LNSKNFNFGGDVVEKLISEQQDLLSKQKYQEAQNLAIFLCDLGNSGVLTAQSIGEYLESF 152

Qy    131 LSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGIQAYLSI 190
          :::|      |:  |  | : |:|:  :|  ||| | || |: |:::| :   |  || :
Db    153 IAAAF-----EENMP--QVRNDYYIQTVLRCLPWIGKELTEKAPEQMENIGEAIGKYLEL 205

Qy    191 RKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLVA 250
          |  :  :|  :  |  :    ::|::| |  :|::| :   |    :|| :  ||  |
Db    206 RNKNHVALLQVWREGSTDQK--QEDYLESLSAQIEALRNADWVENHIPRHYSGFETTLQD 263

Qy    251 GKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRL------NIFPANKMED 304
          | |   :  |  :  |              | :  ||  : ||      : | :  :
Db    264 ALQHNLPSFQSPEHTS--------------DMIYPYPLVVFRLFQDADCSAFSSKPLPG 308

Qy    305 VQPIDRFVVEEYLLDVL----FYLNGCRKECASYM---ANLPVTFRYEYLMAETLFSQIL 357
          ||||: |  :  ::    |    | :|  ::      ::|: |    |: ||:| |:|
Db    309 DSSIDRFLFEGEIAWIIEKNQFNRKACARELLAFAEENPSVPIGF----LIFETIFGQML 364

Qy    358 LLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHH 417
           ||  |:  :::  ::::| |  |  :| ::   |  :: :   :      |:: ||| |
Db    365 RLPHAPYPAIFHCSLVLELLKLKPDDYPQILVQTVECIYRRADSMQPVCIDRMVDWFSFH 424

Qy    418 LSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLP 477
          |||||: : | :|   |:   ::  ::|||:|:::  |   ::||  ::       :: |
Db    425 LSNFQYRYTWTDWKDCLNKDAFSGSQIFVREVIEKCRRFGSYEKIIAALPQ--DFVKIHP 482

Qy    478 PKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWI------EETIYPVHGF 531
              |  | ::|  | |   | :   ::  :|:| |   :  :       :|  | :: |
Db    483 --CSPEVRYLIDE--EDTALVQRAETFTQMFQERQPAEAFLNELKSNDENDELPYNINEF 538

Qy    532 EVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCPDND-KQVMLLSQVSTYWKNNVQ 590
              : |  :| : ||:::|  : | ||       :|  ::  |  ||  : : || | |
Db    539 ----GLFVMVMLKMASKTYSHNFSALFRYQTTLKTVCDASELYQEKLLETLYSCWKTNQQ 594
```

```
Qy     591 MTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPW--EILGNALNKTYNRISDLR 648
           |  :  |:::  :::   |:| |:|  | :.|.|  |: |  |:|  || |   :|: :
Db     595 MLMILTDKLLKMQVIDCSAVVGWLFD-EKMWQEH--DRQWLFEVLNQALEKLTRQINVVE 651

Qy     649 KDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEA 708
           |||  :|:        : |   | :  || : : | |        | :: |  :|
Db     652 KDIKELTEK-------TENKIKEEDDEESDIKMDEDE-------TKEEKFKQDLEDLENN 697

Qy     709 ELSLRESLEAKEALLNRALSETEVLL 734
           :  |   :  :: | |  |   | ||
Db     698 KEKLERMVTFQKGLFNDFLIHAETLL 723


RESULT   7
Q9VDA4
ID   Q9VDA4      PRELIMINARY;      PRT;   747 AA.
AC   Q9VDA4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG7907 PROTEIN.
GN   CG7907.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003734; AAF55893.1; -.
DR   FlyBase; FBgn0038887; CG7907.
DR   InterPro; IPR003890; EIF4G_cent.
DR   Pfam; PF02854; MIF4G; 1.
DR   SMART; SM00543; MIF4G; 1.
SQ   SEQUENCE   747 AA;  85089 MW;  5FC51461A12057E6 CRC64;


  Query Match            12.6%;  Score 550.5;  DB 5;  Length 747;
  Best Local Similarity   25.0%;  Pred. No. 2e-29;
  Matches  167;  Conservative  136;  Mismatches  257;  Indels  107;  Gaps   22;

Qy       7 LLLRIGEKGPEY----GTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHKIP 62
           | |::|  |  |    | | | |:::                  :  |: |    | :
Db      45 LELKLGNLG--YLLGKGMSVDLKENL-----------------INELVICLGYYPGQAS 84

Qy      63 LYGTLIGLLNLENEDF--------VQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLC 114
             | ||:||:|: : :|          |: ||||          : |  | ::  |  |
Db      85 AYATLVGLINVADFEFGSECLLFMAHKVGESVHIR--------DWNRCRGVVHCMVDLYH 136

Qy     115 SKVIQPASLIVVFETLLSSAATTVD-EEKGNPSWQPQADFYVICILSSLPWGGSELAEQV 173
           :|:  : :: :    |       | ::    : | : ||   |:||::|  | :|  :
Db     137 CQVLPSSYILKLLAAFLKDCEALKDPDDLVGVTPQIRRDFLAYCVLSAMPLIGRDLEGET 196

Qy     174 PDEIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGW- 232
           ::::| :| |:  |    |: |: :          ::|::| |  ::  :  | |
Db     197 --AFDKLIVSLQIYIKKRSALHTNMLSVW------PDFNQRDYLELLWQQVDGMRQNHWA 248

Qy     233 --KLESVPRPHLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQR 290
             : : :|||: ||   |  |: |:||           |:  |     |:  |||
Db     249 EPEHQLIPRPYKSFSETLSHGRIHQLR----------DYDLA-----AHEERCRYPLP 291

Qy     291 IRRLNIFPANKMEDVQ------PIDRFVVEEYLLDVLFYLNGCRKECAS----YMANLPV 340
             ||  : : ::          |:| ::| ::||:|   :  || ||     | |: |
Db     292 RVCFRIFSCDSVGEIPNMPPPVSIERHLLEAHILDILISFHKERKICADSLLMYAASKPQ 351

Qy     341 TFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKIS 400
           | | : | :  ::| ||   : |: |  ::::|||  |   | |||  |:  :: :::
Db     352 LPVY-YCIVEVILGEMLRLPTANWSTIAYGSILVELCKRQPDKIPQVVAQALDIIYNRLN 410

Qy     401 DLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLP------KWAPKRVFVQEILQREV 454
           : :    ||: | |||:||| |   | :||  |  |         || ||::|:|::
Db     411 SMSVACFDRLVNWVSHHISNFGFNCQWSKWAQGLPSPIDPSATNLQPKVVFLRELLKKCF 470

Qy     455 RLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTA 514
           |      ||  : :   | : |||   |:| :      |      || :|   ::  | :
Db     471 R------IKDVVPD--ILADFLPPPVLPHFKFV----DETLPGAILSKDLLEAMRSPQAS 518

Qy     515 RDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKL-CPDNDK 573
           :||  | ::   : |  : : :  |  | :|||||:|   :| :|  ||  |  | ::
Db     519 PEMISEIIKSSTGI-GPLLKINVFTQNCLHLGSKSFSHTFGILRKYHSVFKDLVAGDPER 577

Qy     574 QVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSP-.--ENVDQFHVSDQPW 630
           |  :|: :   |  : |   |  :::   ::   ::|||:| |   : : : ::   |
Db     578 QAAVLNGIFDVWVASDQYKFVVTEKLVTGLIIEPISVVRWIFGPSMRKELTKIYI----W 633

Qy     631 EILGNAL 637
           |:| :||
Db     634 ELLHSAL 640


RESULT   8
Q9HE91
ID   Q9HE91      PRELIMINARY;      PRT;   829 AA.
AC   Q9HE91;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   RELATED TO CAP BINDING PROTEIN 80 (CBP80).
GN   B13O20.110.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
```

```
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL451015; CAC18201.1; -.
SQ   SEQUENCE   829 AA;  92651 MW;  BFBABCBB76304ACB CRC64;
```

```
  Query Match             11.5%;  Score 501.5;  DB 3;  Length 829;
  Best Local Similarity   21.6%;  Pred. No. 5.6e-26;
  Matches  176;  Conservative  156;  Mismatches  319;  Indels  165;  Gaps   26;

Qy      50 LLQCAEQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFM 109
           :|| | : | | |    :: : |      |  |:  : :  :  :  |:   ::: |:
Db      86 VLQLAVEQPMKTPFLAAVVLVANTNKPQIVDMLLAKLTSTLETKIAEGDWRDVKLYLKLA 145

Qy     110 TSLLCSKVIQPASLIVVFETLLS--------SAATTVDEEKGNPSWQPQADFYVICILSS 161
           |     ::   :  | | | |        |:  |:  |              |  || :
Db     146 GCL--QSCLEGDGVFPVLEELFSRAVDLQTASSEDTIGTE-----------LVKIILLT 191

Qy     162 LPW----GGSELAEQVPDEIER--VLVG----IQAYLSIRKNSSTSGLNFFHNGEFESSL 211
           ||:    | ::  ::  | :|:  :: |    :|| :          :     :| :
Db     192 LPYVMVTGPTQWQQKAADLMEKTEIIAGEAHPLQALIE----------PYHPEAGDQSPV 241

Qy     212 AEKDFVEDLLDRIQSLASNGWKLESVPRP------HLSFEAQLVAGKFHELRPIKCME-- 263
           | :  :  |  ::|: |:||| |  :|||       :  : ::   : | | ||   :
Db     242 ASQSCISLLQTQLQNEANNGWALSCLPRPWELPVEEVEQQEKIANAEKHTLPPISIPQTV 301

Qy     264 ----QPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPPANKMEDVQPIDRF---VVEEY 316
               :|  |  :   |: :                    :|   |:      :: :
Db     302 VAGPRPLFPEIYFSVYANQD---------------------VESTPPLTNIAASLIRDA 339

Qy     317 LLDVLFYLNGCRKECASYMANL-----PVTF-------------------RYEYLMAET 351
           |:| :  ||  |   | ::  |     | ||                  : | :  :
Db     340 LVDTINVLNFNRNVTARFLIELDCYFAPRTFAARATPFDKIREIEPPKSSWKPEDVAVDA 399

Qy     352 LFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLI 411
           :|||:  || |  | :||  |: : ||  | |     :  |:| |:     :|:|   | :
Db     400 VFSQLFSLPNPEHKLVYYHAVLTEACKLAPAAIAPSLGRAIRFLYRNSHRMDLELSNRFM 459

Qy     412 LWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATA 471
           |||||||| | | | ||   : |   | : |:   : :|:|||:  :|| ::
Db     460 DWFSHHLSNFGFTWKWTEWVDNVYLSNLHPDKAFIMGAIDKEIRLSFAQRIKNTLPE--P 517

Qy     472 LEELLPP---KAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPV 528
           : |: |   |  |:| :: ::    | ::::| | ||  |::       ||  |  :
Db     518 YQSLIGPEKEKDVPDFKFADDKTPFAAEGKEIAALLRRKAPEEE--------IEPVIERI 569

Qy     529 HGFEV---------TLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCPDND-KQVMLL 578
           |   :          :  : | ::| :|||| :|::  :||  :  :      ::  :  ::
Db     570 HSLALDNNLDPLVASTDVFVTSVLHVGSKSLSHVLAAIERTKERLTDAGATSEAARTQII 629

Qy     579 SQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALN 638
           | |  ||  :  :    |:::: | ::: ||:: |           ::       | ||
Db     630 SSVMEYWSAHPGVAIAIIEKLLNYSILTPQAVINWA----------ITTYAGATRGEALA 679

Qy     639 KTYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRL 698
           | :          :  | :|  | ::  |  |: |    :::: | :  |    |:  |
Db     680 KGF---------VYEMVFNTVV--KVTSRLRQVLQKATLPEAMIDDETIEAEING-MRSL 727

Qy     699 KSTVEKT--GEAELSLRESLEAKEAL-LNRALSETEVLLLLLFQSFLGVLKER------- 748
             :|      |  |  | ||| : |      |||| |:    : :| | | |
Db     728 FRAIEDALFAWASGSKDEMLEASDGLGEGDGTSETEKLVKRWGERWLRVFKRRAAIEEAF 787

Qy     749 LPDPTKVR------SVQDLKSIGAEDDKPSAMDVDS 778
           : :  | |      :  :| :  |::|  :   ||:
Db     788 IVEARKERERKAAAAAAELAAAPAQEDDMAVDSVDA 823
```

```
RESULT   9
O14253
```

```
ID   O14253      PRELIMINARY;      PRT;   780 AA.
AC   O14253;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 90.1 KDA PROTEIN C6G10.07 IN CHROMOSOME I.
GN   SPAC6G10.07.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Oliver K., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO YEAST GCR3.
DR   EMBL; Z98603; CAB11293.1; -.
DR   InterPro; IPR003890; EIF4G_cent.
DR   InterPro; IPR002173; PfkB.
DR   Pfam; PF02854; MIF4G; 1.
DR   SMART; SM00543; MIF4G; 1.
DR   PROSITE; PS00584; PFKB_KINASES_2; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   780 AA;  90077 MW;  6AAA6562287F74F7 CRC64;
```

```
  Query Match             10.9%;  Score 475.5;  DB 3;  Length 780;
  Best Local Similarity   23.9%;  Pred. No. 3.2e-24;
  Matches  192;  Conservative  149;  Mismatches  299;  Indels  163;  Gaps   36;

Qy      40 ERSGDQVLPFLLQCAEQ----LPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALD 95
           | :  : :  :| | :     :| |||   |||   :|     |  ::|   | | :
Db      67 EANDPETISTILDCIQTTAFIIPVKIPHLATLIIRASLR----VPLILEKAAAYFCLQYF 122

Qy      96 SGNCNSI-----RILLRFMTSLLC-SKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQP 149
           : | ||        :: ||    |:| |  :|| :|    : | |  |  : :|     |
Db     123 T-NLNSFLYYEAKVDLRM---LICMSFALQPGTL----KPLFSLLADAISKETKPSVW-- 172

Qy     150 QADFYVICILSSLPW--------GGSELAEQVPDEIERVLVGIQAYLSIRKNSST--SGL 199
             | ::  || :||:          |  : | :: |: |        |:  ||:| |  : |
Db     173 -GDNFLRIILINLPYFIAANNDLGKKDFANEILDQCE-------IYVRHRKSSITLSNPL 224

Qy     200 NFFHNGEFESSLAEKDFVEDLLDRIQSLA-SNGWKLESVPRPHLSFEAQLVAGKFHELRP 258
           :   |       |:|::   ||| :   |:  | :    : :|   ||:  |     | :
Db     225 SIHDN------LSEEEL--DLLYKQLILSRENDFTFPYISQPWKFFESDFV----HIV-- 270

Qy     259 IKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPPANKMEDVQPIDRFVVEEYLL 318
               |  ||      :  : | : :      | |       |      :  :  : :
Db     271 ------PVSPSIPEWTFQPTPQQNELPSFKRFFELFNNFEIRTTPDASDVAASIFRDISV 324

Qy     319 DVLFYLNGCRKECASYMANLPVTFRY------------------------EYLMAETLF 353
           ||: :|   | | |  : :| | | |                        | :: | :
Db     325 DVINHLEFNRVEAAQVLTDLDVYFTYKTFALRGTPVNELPNLDPSESRWKAEDIIVEAVL 384

Qy     354 SQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILW 413
           ::|      :| :||  :::: |:  |          :| ::   ||| :::  | | |
Db     385 GELLGSQNTTYKPVYYHSLLIECCRIAPKILAPTFGRVIRLMYTMSSDLPLQTLDRFIDW 444

Qy     414 FSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALE 473
           |||||||| | | | ||    ::|    ||:||::| : ||: |||: :|  |:      |
Db     445 FSHHLSNFNFHWKWNEWIPDVELDDLHPKKVFMRETITRELILSYYTRISDSLP-----E 499

Qy     474 ELL----PPKAGPNFMYSLEEGKEKTEEQQL--SAELSRKVKEKQTARDMIVWIEETIYP 527
           ||        :||||:|  |      :  |:  :  | : ::|     |:|:  ||
Db     500 ELRCLLGEQPSGPNFVYENETHPLYQQSSQIIEALRLHKPLEE----LDIILQSEE---- 551

Qy     528 VHGFEVT-LTIVVQTLLDIGSKSFTHLVTVLERYGQV---FSKLCPDNDKQVMLLSQVST 583
           :   | : : :|:     :||:||:| : | |::      ||:     | :: :: :: :
Db     552 IQNSETSAVRLVMSCAYSLGSRSFSHALNVFEKHLNTLKHFSR--KSLDSEIEVVDELFS 609

Qy     584 YWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNR 643
```

```
Db    610 FWKLQPFNAVMWLDKMLNYSIISITSIIEWLIK-QDVTIWSRS-YTWSL----VNTTFNK 663

Qy    644 I-SDLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTV 702
Db    664 LAARLRRSVSNKEDSSLINE-----ANEEKEIVTNLL--------------LSALRALI 703

Qy    703 EKTGEAELSLRESLEAKEALLNRALSETEVLLLLLFQS-FLGVLKERLPDPTKVRSVQDL 761
Db    704 SENAE---------------NIWVSHWLNLMLKYVESNFLSVKKDTIEEANE--PVQEN 745

Qy    762 KSIGAEDDKPSAMD-VD---SEN 780
Db    746 TSEEQEDTKMQPVDAVDEQPSEN 768


RESULT 10
Q9Y717
ID   Q9Y717      PRELIMINARY;      PRT;   886 AA.
AC   Q9Y717;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CAGCR3 PROTEIN.
GN   CAGCR3.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IFO1060;
RX   MEDLINE=99221032; PubMed=10206191;
RA   Uemura H., Nakamoto K., Sugioka S., Tadenuma M.;
RT   "Isolation and sequence of the GCR3 homologue from Candida albicans by
RT   complementation of delta gcr3 strain of Saccharomyces cerevisiae.";
RL   Yeast 15:323-327(1999).
DR   EMBL; AB015281; BAA76378.1; -.
DR   InterPro; IPR003890; EIF4G_cent.
DR   Pfam; PF02854; MIF4G; 1.
DR   SMART; SM00543; MIF4G; 1.
SQ   SEQUENCE   886 AA;  103127 MW;  BEB223BDC8864B4E CRC64;


  Query Match             8.6%;  Score 374.5;  DB 3;  Length 886;
  Best Local Similarity   21.5%;  Pred. No. 3.5e-17;
  Matches  191;  Conservative  143;  Mismatches  313;  Indels  241;  Gaps   35;

Qy     10 RIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHKIPLYGTLIG 69
Db     51 RLGEAGDIENLIAD-TNYISNPIVAEFEKIDNLRNSILSTIYALITEQPHKISAIANLIL 109

Qy     70 LLNLENEDFVQKLVESVHANFQVALDSGN---CNSIRILLRFMTSLLCSKVIQPASLIVV 126
Db    110 ICNAKNFVIAKYVIEYLHSKMQTMLDSQEGSVFNDIKNILKFLSTL--TPIIEDNGIIQI 167

Qy    127 FETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIER---VLVG 183
Db    168 FKQFL-NFAIELQEQTEVRNGLAQEIYYNVLI--AIPY---VLSNDNSDDLKKSINELIE 221

Qy    184 IQAYLSIRKNSSTSGLNF-FHNGEFESSLAEKDFVEDLLDRIQSLASNGWK--------- 233
Db    222 LARNFKIVAESATVLLPFDTRMGNFELPYIPKQMVDLILPALIKLQNDDWNFRLFLDFKT 281

Qy    234 -LESVPRPHL---SFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGK---------- 278
Db    282 HLDPVIQSALENNSISSELVKHKLPQL---------SFPSVENASFKSTSIDKLWNDNPR 332

Qy    279 ---QKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFV---VEEYLLDVLFYLNGCRKECA 332
Db    333 YLFQVYNNTTEY--------------ETVPPIETYLGLFFKDIAFDILTNLSFNKNETA 377

Qy    333 SYMANLPV--------------------------------TFRYEYLMAETLFSQILL 358
```

```
Db    378 IELSILDMFFDNKLFAPPGTSIDQLNAIYEDNKSGTNDPSLSTWKIEDVAVESILTMIFQ 437

Qy    359 LPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHL 418
Db    438 LPNPLEVEIYYYTVLISCCRESPESIAPVFGRAIRYFYNHLETLDYELKIRFLDWMSIQL 497

Qy    419 SNFQFIWPWEEWAF----VLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSI-----ENA 469
Db    498 SNFEFSWKWEEWVSDSKKLKDL-KYHPKKNFIKNLIAKEIRLSNKKRIKDSFIDIVDGEV 556

Qy    470 TALEE--------LLPPKAGPNFMYSLEEGKEKTEEQ--QLSAELSRKVKEKQT--ARDM 517
Db    557 VNLDEFYQYLDISMVPDVKSYIISYDTELYGESSRDTLIQLYEQKQEQLNSKSNIGAQNE 616

Qy    518 IVW--------IEETIYPVHGFEVT---------------------------LTIVVQT 541
Db    617 IFFNFTNSELPFHETASKVYDFILTHWKSNTDFNELYKSVLESITAPNNERFAINLIIQT 676

Qy    542 LLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTY---------------WK 586
Db    677 YAYIGSRSIYSVVSILSR----------DINKLKFLSGAPIDYVGDEARFEDLHFTEEEK 726

Qy    587 NNVQMTAVAIDRMMGYRLVSNQAIVR-WVFSP-------ENVDQFHVSDQPWEILGNALN 638
Db    727 QNRQNWII-------------EAVFRIWIHQPQVVFLILEYLIEFGIID-PKYILVKALE 772

Qy    639 KTYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRL 698
Db    773 S--NLIID----------NVSCMESIN---RI-LSKAESKELIIQ------------LF 803

Qy    699 KSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLFQSFLGVLK 746
Db    804 TAIIDNLNKLELDKEKRVEIEEVTESNA---TEVDKQWLFYEYLGLLK 848


RESULT 11
Q9XWR0
ID   Q9XWR0      PRELIMINARY;      PRT;  1413 AA.
AC   Q9XWR0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Y11D7A.14 PROTEIN.
GN   Y11D7A.14.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Steward C.A.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; AL032632; CAA21588.1; -.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00242; MYSc; 1.
SQ   SEQUENCE   1413 AA;  163882 MW;  0ED5080C8A6E5B6A CRC64;


  Query Match             3.7%;  Score 163;  DB 5;  Length 1413;
  Best Local Similarity   18.1%;  Pred. No. 0.026;
```

```
  Matches  160;  Conservative  141;  Mismatches  301;  Indels  284;  Gaps   39;

Qy      13 EKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFL-LQCAEQLPHKIPLYGTLIGLL 71
           |:| :: |  :: :|::    :| : :      :| ||  :|   :|:     |:   ||
Db     440 EEGIKW-TQVNFANHLQPTIDLIEKPM-----GILSFLEEECV--VPN-----GSEKSLL 486

Qy      72 -----NLENED----------------FVQKLVESVHANFQVALDSGNCN---SIRILL 106
                || |:                 :||     || |   ||    |   |:  :|
Db     487 EKLCSNLSNDSSFKKSKQTQKCSTIRHFYVQHYAGEVHYNIDGWLDKNRDNVETSVLDIL 546

Qy     107 RFMTSLLCSKVIQPA----------------------SLIVVFETLLSSAATTVDEEKG 143
             |  |   :  |                        |  : :|| ||:|
Db     547 SQSTHPLLKLLFPPVPVNNLKTRRGTITNSTVSFLYKNQLQCLLDTLNSSSA-------- 598

Qy     144 NPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFFH 203
                 :: |::|:         |::| :|:  ||  |    :: |    |:
Db     599 ---------HFIRCVVSNY--------EKLPGKIDAPLVLAQ----LKCNGVLEGIRICR 637

Qy     204 NGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELRPIKCME 263
           | : | |:  :|:|    |   |  |
Db     638 EG-YPSRLSHSEFIE----RYSLLMKN------------------------------- 659

Qy     264 QPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFY 323
             : |:  | |:|   :                :| |
Db     660 -----KEQSKGASEKEKCTLI----------------CQDAQ---------------- 680

Qy     324 LNGCRKECASYMANLPVTFRYEYLMAET-LFSQILLLPQPPFKTLYYTLVIMDLCKALPG 382
             |||               | : :| || :: :: :   |   |
Db     681 ---VRKE-------------RYAVGKTKLFCKVGVISELETKRNNYI----------S 712

Qy     383 AFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHLSNFQFI--WPWEEWAFVLD--LPK 438
           :|  ::   :| | :  ::   :  ::    ::  |  :  |||    :   :|:
Db     713 SFIILIQANIRYLNIQKDLIERRKKLEAVVTIQDNVRQFAELSQWPWYRIYHLTRGLIPR 772

Qy     439 WAPKRVFVQEILQREVRLSYWDKIKQ-SIENATALEELLPPKAGPNFMYSLEEGKEKTEE 497
             |   ::|:   :::|   ::|::  |:|  ||:|       |   |:   :||:|:
Db     773 NRDKE-RIEELENEKLKLE--EEIQEMEIKNEEALKE--------NLKLSMLLDREKSEK 821

Qy     498 QQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVL 557
            ::  ||    |: :         |: :     |  |:  : }            :  ||
Db     822 VKVQKELEEVEKQGR---------EKLLEKEREFRKTMEEMEQN---------EEIFNVL 863

Qy     558 ER-YGQVFSKLCPDND------KQVMLLSQVSTYWKNNVQMTAVAIDRM-MGYRLVSNQA 609
           || | :   |:  ||      :::  |:   |  :|  |       :|    | :  :
Db     864 ERKYNEQHKKVMKMNDVLREYERKIEQLNMEKTDLENENQKLRETQNRQDSHYSNLEKEV 923

Qy     610 IVRWVFSPENVDQFH-VSDQPWE------ILGNALN-------KTYNRISDLRKDISNIT 655
           : :    |  :|   :||:  |        |  ||        :  | | |::| || :
Db     924 MEKSSLIDELQNQIQKLSDENNEQRLTIAKLETALEDEKARFARQNNTIGDMQKLISELN 983

Qy     656 KNVLVAEKASANAR-----VELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAEL 710
           : :   :  : | |     :| |  :    |   : :: :   |:  ||    |
Db     984 EKIARFDNIALNERNSTRKIEREKEKLNEELTTAKEIIQKQAKKIDELKEECRKRKNEAS 1043

Qy     711 SLRESLEAKEALLNRALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDK 770
             |  || |||::   : | :                |||| :    : |:|:|   ::
Db    1044 RLERKLEDKEAMMADCVKELK----------DSHKERLKE--MEQKVEDVKRKNSK--- 1087

Qy     771 PSAMDVDSENGNPKKSCEVGERE-----QWCLSTLGYLTAFTRQYA 811
             :::||   |   |  :||       :  |: | |:   |||:
Db    1088 -----LENENSTQKSQIETFQRESSVDSDYGRSSSGRLSTLGRQYS 1128


RESULT  12
Q9VM67
ID   Q9VM67      PRELIMINARY;      PRT;  1833 AA.
AC   Q9VM67; Q9VM66;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG18304 PROTEIN.
GN   CG18304 OR CG13774.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003615; AAF52458.2; -.
DR   FlyBase; FBgn0031869; CG18304.
DR   InterPro; IPR002017; Spectrin.
KW   Hypothetical protein.
SQ   SEQUENCE   1833 AA;  206010 MW;  ED1948482573584E CRC64;


  Query Match             3.3%;  Score 145.5;  DB 5;  Length 1833;
  Best Local Similarity   19.7%;  Pred. No. 0.61;
  Matches  177;  Conservative  112;  Mismatches  319;  Indels  291;  Gaps   32;

Qy     118 IQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPD-- 175
           : |  : |   |  ||: ::   | : |     :  |:   ||     | :  :  |
Db     235 LPPKKVAVASTTTASSSNSSSTSLKTSNSTSASNEVKVVTSTSSSSTSSSSVRRKEADSV 294

Qy     176 ---EIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGW 232
              ||:|  |      ||  ::|||           :| | |           |  :||
Db     295 ASKEIKRQTV---PAASISHSNSTSS---------TASTASK----------SQDTNGM 331

Qy     233 KLESVPRPHLSFEAQLVAGKFHE-------LRPIKCMEQPSPPSDHSRAYSGKQK----H 281
           : | :    |  |       :  |       || :  |:  |  :  | | : :||
Db     332 Q-EQMKALKLELETMKTRAEKAEREKSDILLRRLASMDTASNRTAASEALNLQQKLNEMK 390

Qy     282 DALTRYPQRIRRLNIFP---ANKMEDVQPIDRFVVEEYLLDVLFYLN-GCRKECASYMAN 337
           : | |  :  |:||:      ||  : :   :    | : : |   |   :|| :  |
Db     391 EQLDRVTEDKRKLNLRMKELENKGSESELRRKLQAAEQICEELMEENQSAKKEILNLQAE 450
```

```
Qy    338 LPV---TFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRA 394
          :     ||| : : |:|   :                  || ||       :::  ::
Db    451 MDEVQDTFRDDEVKAKTSLQK-----------------DLEKATKNC--RILSFKLKK 489

Qy    395 LFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREV 454
            ||  |: | ::     |:  |||                            :| : |
Db    490 SDRKIETLEQERQSS----FNAELSN-------------------------KIKKLEE 518

Qy    455 RLSYWDKIKQSIENATALEELLPP--KAGPNF-------------MYSLEEGKEKTEEQ 498
          | : ::: : ::     |||  |  |  |                 ||  |  : : |
Db    519 ELRFSNELTRKLQ--AEAEELRNPGKKKAPMLGVLGKSTSADAKFTRESLTRGGSQEDPQ 576

Qy    499 QLSAELSRKVKEKQTARDMIVWIEETIYPVHG-------------FEVTLTIV------- 538
          |  ||   :: :   :| : : || :  :                 || | :|
Db    577 HLQRELQDSIERETDLKDQLKFAEEELQRLRDRERKRVRFSCGTQTEVPLEVVAFPRGTQ 636

Qy    539 -VQTLLDIGSKSFTHLVT---------------------------------VLERYGQ 562
           | |:    | |  :|||                                  | ||
Db    637 TVATVQSDMSTSVENLVTSNVAVTQTDFEVPDRNVSIERETMSSPFAGLFPPSSSSRVGQ 696

Qy    563 VFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVF----SPE 618
            || |      :                |  ||  |     : | :| :
Db    697 SGRKLSPTPHPHRL-------------APEVHADRDEGISDEDDPAELRILLELNEQEA 742

Qy    619 NVDQFHVSDQPWEILGNALNKTYNR--ISDLRKDISNITKNVLVAEKASANARVE----- 671
          :: :  | |   |   || :| | |   : ||:| || :|: |::   |::|  :
Db    743 SILRLKVEDLEKE---NAESKKYVRELQAKLRQDSSNGSKSSLLSLGTSSSAAEKKVKTL 799

Qy    672 -------------------LEAAESKLSLVEGE-PVLGENPAKMKRLKSTVEKTGEAEL 710
                            |:  |||  :| |   | :   ::  |:  |||||  :
Db    800 NEELVQLRRTLTEKEQTVDSLKNQLSKLDTLETENDKLAKENKRLLALRKASEKTGEVDQ 859

Qy    711 SLRES-------------------LEAKEAL----------------------LNRALS 728
           ::||                  |||:: |                      |   :|
Db    860 KMKESLAQAQRERDELTARLKRMQLEAEDKLPPRTAKRVNDLTPKSHLKKWVEELEDEIS 919

Qy    729 ETEVLL-------LLLFQSFLGVLKERLPDPTKVRS-----VQDLKSIGAEDDKPSAMDV 776
          |  |:|       |   || | |:| |     :  |     || || :     | | ::
Db    920 EMRVMLSSSGTDQLKALQSAKGALEEDLRKCKQKLSLAEGDVQRLKLLNGSSSKVSELEQ 979

Qy    777 DSENGNPK--------KSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEVFS 827
            : |: :        |  |   ::|     ||       : :   |    | | |  |
Db    980 KLKRGDEEAKKLNSKLKDLEDKVKKQEAQLKLG-------ETSKSTWESQSKREKEKLS 1031
```

```
RESULT  13
Q92B64
ID   Q92B64      PRELIMINARY;      PRT;  1023 AA.
AC   Q92B64;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SBCC PROTEIN.
GN   SBCC OR LIN1686.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / SEROVAR 6A;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596169; CAC96917.1; -.
DR   ListiList; LIN01686; -.
KW   Complete proteome.
SQ   SEQUENCE   1023 AA;  116266 MW;  A60D7A03C8A80B5E CRC64;
```

```
  Query Match             3.3%;  Score 145;  DB 16;  Length 1023;
  Best Local Similarity   16.8%;  Pred. No. 0.28;
  Matches  155;  Conservative  159;  Mismatches  297;  Indels  312;  Gaps   38;

Qy      6 TLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIER---SGDQVLPFLLQCAEQLPHKIP 62
          ||: |:             || |      : ::|| :   :|    |   :  | :  ::
Db     80 TLVFRL------------KDKIYQISRIPQQEIAKQRGNGTTTSPQKAELYELIGDEMK 126

Qy     63 LYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRF----MTSLLCSKVI 118
          |  : :  :| : |: :|                  | :  | :|         || |
Db    127 LLASSVSDVNTKMEELIQL---------------NVDQFRQILMIPQGEFRELLVSDSK 170

Qy    119 QPASLIVVFETLLSSAATTVDEEK-GNPSWQPQADFYVICI--------LSSLPWGGSEL 169
          :       | :|   | ||  ||  |  |: | :  :: :          |: |   | |:
Db    171 EK-------EVILQRLAHTVYYEKVENLLWEKQKEAEILVVEARKKVAELAELSLPGIEI 223

Qy    170 AEQVPDEI-----------ERVLVGIQAYLS-IRKNSSTS--------------GLNFF 202
            :  ||             : :|  ::: || ||| :| :                |: :
Db    224 TGKTTTEISLLQTEAIKQEQAILAELESALSIIRKETSEAVEKVTLAKEQLLDWQNLDLY 283

Qy    203 HNGEFESSLAEKDFVEDLLDRIQSL--ASN----------GWKLESV--PRPHLSFEAQ 247
            |     |||| : : :||::   |||              :||:       :: | :
Db    284 IE-EVAKLEAEKDFYQVIANRIEAAKRASNLRSQDALCIRLKEQLETAVNTEKQVAHEVE 342

Qy    248 LVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANK-----M 302
          |:  :|                    : |:: :||       :  |||      :
Db    343 LITNQFS-------------------NAKKQKEALAEQEA------VLEANKRTLFQL 375

Qy    303 EDVQP----IDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILL 358
          |:::|    ::   :::   :: :      || |          | | :|
Db    376 EEMEPKIIELETVTIQKRRAEINW------KEAAD---------RLEKVM---------- 410

Qy    359 LPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHL 418
                ||                    :|  ::::  :: :::
Db    411 ------KT-----------------EQEIAAELQSMETRLGEINQ-------------- 432

Qy    419 SNFQFIWPWEEWAFVLDLPKWAPKRVFV---QEILQREVRLSYWDKIKQSIENATALEEL 475
                  | | :  : |       :   ||:: : :::  |:| ||: |    |::|
Db    433 ---------AELAILEAINKRTTAEALIEKNQELVNKRMKMDLWNKQKQTEEQ--TLKQL 481

Qy    476 LPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTL 535
          |  |     |   |   :|: :| |:  : |                  || |
Db    482 LTEKLAIETMIKQE---ESNLQQEQAATLAAHLHEGDAC-----------PVCG----- 521

Qy    536 TIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCPDN--DKQVMLLSQVSTYWKNNVQMTA 593
          ::    | : |  :   | |: |   ::  |    |  :| :  |       | :  :
Db    522 SVSHPQLAEYGESA--SLTTLEEAKAKLHEKQLAINEVEKSISQLEWQLAEWADIADVDL 579

Qy    594 VAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISN 653
          ||::::    |  |: : : :    |  | :|: :  | :  ||    |:  :|  : :|
Db    580 VAVEKL----LAENRQLAKNLTDQINQLQTNVAQK--ENIQGALETLKNKQKELETEKNN 633

Qy    654 ITKNVLVAEKASANARVELEAAESKLSLVEG---------------------------- 684
              |       :   |:: :  ||| :|
Db    634 TALQV-------ESFHQEVQLSSGKLSYLEQAIPADLRDKTIFDKKKNELSNSIKTHLEQ 686

Qy    685 ----EPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLFQS 740
              : :  |   :  ||:||::      |  :|:|: :  :   :: : :       | :
Db    687 AEQVDKIFREAEKETTRLESTLQSAENTTLDAKEALQVQREVFKESMKQND------FPT 740

Qy    741 FLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTL 800
          :    |: |       ||::|||                      |  :: : |:     :
Db    741 Y-EAYKQAL------MSVEELKS--------------------KEEQLADFERKRHLAI 772

Qy    801 GYLTAFTRQYASEIWPHMEKLES 823
```

```
                 |  :   ::  |::|:|||
Db    773 SRQADLTEKLKNKQKPNIEQLES 795


RESULT  14
Q96JN1
ID   Q96JN1      PRELIMINARY;      PRT;   802 AA.
AC   Q96JN1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA1794 PROTEIN (FRAGMENT).
GN   KIAA1794.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=21245130; PubMed=11347906;
RA   Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XX.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large Proteins in vitro.";
RL   DNA Res. 8:85-95(2001).
DR   EMBL; AB058697; BAB47423.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   802 AA;  90592 MW;  812DAF16CB7DA6B1 CRC64;


  Query Match             3.3%;  Score 142.5;  DB 4;  Length 802;
  Best Local Similarity   20.0%;  Pred. No. 0.29;
  Matches  164;  Conservative  132;  Mismatches  289;  Indels  237;  Gaps   38;

Qy      72 NLENEDFVQKLVESV------HANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLIV 125
           :: || |  ::::|:       |: :: |  |  : :|       | | : |:|
Db      51 SVANETFCLEIMDSLRRCLSQQADVRLMLYEGFYDVLR-----RNSQLANSVMQ------ 99

Qy     126 VFETLLSSAATTVDEEKGNPSWQPQADF-----YVICILSSLPWGGSELAEQVPDEIERV 180
             ||||              ::|: |         |||:     | ::: | |  :: :
Db     100 ---TLLSQL---------KQFYEPKPDLLPPLKLEACILTQ----GDKISLQEP--LDYL 141

Qy     181 LVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRP 240
           |  ||  |:  ||:          || |    |: | ||| | ::|: :   | |
Db     142 LCCIQHCLAWYKNTVIP----LQQGE-EEEEEEEAFYEDLDDILESITNRMIKSELE--- 193

Qy     241 HLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKH----------DALTRYPQ 289
           ||    |                  | |:: |   |:           : |  |
Db     194 --DFELDKSA----------------DFSQSTSIGIKNNISAFLVMGVCEVLIEYNF 232

Qy     290 RIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNG-CRKECASYMANLPVTFRYEYLM 348
           |   : |  |: ||:  :  |:  : | |:|    |  : : |:  ::  :: ::   :
Db     233 SI---SSFSKNRFEDI--LSLFMCYKKLSDILNEKAGKAKTKMANKTSDSLLSMKFVSSL 287

Qy     349 AETLFSQILLLPQPPFKTL--------YYTLVIMDLCKAL---------PGAFPAVVAGA 391
           ||   :  |     |        |  | :  : |          |  |
Db     288 LTALFRDSIQSHQESLSVLRSSNEFMRYAVNVALQKVQQLKETGHVSGPDGQNP------ 341

Qy     392 VRALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQ 451
           :|: : ||     ||::||                          ::      |:|  :
Db     342 -EKIFQNLCDL-----TRVLLW----------------------RYTSIPTSVEESGK 371

Qy     452 REVRLSYWDKIKQSIENA-TALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKE 510
           :|  |      : ::   :|:::  ||    |: :|:   :: ||::  |::|
Db     372 KEKGKSISLLCLEGLQKIFSAVQQFYQPKI-QQFLRALDVTDKEGEERE-DADVS----- 424

Qy     511 KQTARDMIVWIEETIYPVHGFEVT-LTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCP 569
                 : | : :  |: : | ::     |  ||    |||||     ::   | |
Db     425 ---------VTQRTAFQIRQFQRSLLNLLSSQEEDFNSKEALLLVTVLTSLSKL---LEP 472

Qy     570 DNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVS-DQ 628
            : : | :||  |   | | :  |:    :|            :||       |||
Db     473 SSPQFVQMLSWTSKICKENSREDALFCKSLMNL----------LFS------LHVSYKS 515

Qy     629 PWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLV---EGE 685
           |  :| :     :  : |: :|:          || :  | | |  |   : |:   : |
Db     516 PVILLRDLSQDIHGHLGDIDQDVE--------VEKTNHFAIVNLRTAAPTVCLLVLSQAE 567

Qy     686 PVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLL-LLFQSFLGV 744
            || |    : :||  |     :: :| |   ::  | |: : :  :: |  |   | :
Db     568 KVLEEVDWLITKLKGQV-----SQETLSEEASSQATLPNQPVEKAIIMQLGTLLTFFHEL 622

Qy     745 LKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLT 804
           ::  ||  : | ::  || :                    |::           ||
Db     623 VQTALPSGSCVDTL--LKDL----------------------CKM----------YTTLT 648

Qy     805 AFTRQY------ASEIWPHMEKLESEVFSGEDVHPLFLQAIS 840
           |  | |      :  |  :||||     ||  : ||    ||
Db     649 ALVRYYLQVCQSSGGIPKNMEKLVK--LSGSHLTPLCYSFIS 688


RESULT  15
Q9QZ84
ID   Q9QZ84      PRELIMINARY;      PRT;  2473 AA.
AC   Q9QZ84;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LEK1 (FRAGMENT).
GN   6530404A22RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99303627; PubMed=10373470;
RA   Goodwin R.L., Pabon-Pena L.M., Foster G.C., Bader D.;
RT   "The cloning and analysis of LEK1 identifies variations in the
RT   LEK/centromere protein F/mitosin gene family.";
RL   J. Biol. Chem. 274:18597-18604(1999).
DR   EMBL; AF194970; AAF07196.1; -.
DR   MGD; MGI:1915046; 6530404A22Rik.
FT   NON_TER       1       1
SQ   SEQUENCE   2473 AA;  281903 MW;  504E91A1A150A3E5 CRC64;


  Query Match             3.2%;  Score 142;  DB 11;  Length 2473;
  Best Local Similarity   19.6%;  Pred. No. 1.6;
  Matches  195;  Conservative  121;  Mismatches  321;  Indels  356;  Gaps   41;

Qy      33 GVIRREIERSGDQVLPFLLQCAEQLPHKIPLYGTL-------------------IGLLN 72
           | :  |::   | :          |    |  |  |                   || ||
Db    1479 GAVESEVKDKADLIQTLSFNVGELTKDKAHLQEQLQNLQNDSQELSLAIGELEIQIGQLN 1538

Qy      73 LENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLIVVFETLLS 132
           | |  |::       |||: |    |                              :  :|
Db    1539 KEKESLVKE-----SQNFQIKLTESECE-------------------------KQTIS 1566

Qy     133 SAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGIQAYLSIRK 192
           |     :|||  : |          |||       |   |:   ||:: | |:|    :|
Db    1567 KALEVALKEKGEFAVQ----------LSS----AQEEVHQLRRGIEKLSVRIEA--DEKK 1610

Qy     193 NSSTSGLNFFHNGEFESSLAEKDFVED---LLDRIQSLASNGWKLESVPRPHLSFEAQ-L 248
           : |          : :  | | ::|    |:|     :|  :   :|  | : |
Db    1611 HLSAV-------AKLKESQRESDSLKDTVETLERELERSEENQELAILDSENLKAEVETL 1663

Qy     249 VAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQ------RIRRLNIFPANKM 302
           | |    : ::  |        |    || :  :|  :        :|||       :
Db    1664 KAQKDEMTKSLRIFELDLVTVRTERENLAKQLQEKQSRVSELDERCSSLRRL-------L 1716

Qy     303 EDVQPIDRFVVEE----YLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILL 358
           |: :   |  :||     :| :   |   |:| |:   :            ||| :|
Db    1717 EEKEQA-RVQMEEDSKSAMLMLQMQLKELREEVAALCND----------QETLKAQEQS 1764
```

```
Qy     359 LPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWF---S 415
           | |                  ||     :  ::| |   | | | :    ::
Db    1765 LDQ------------------PGEEVHHLKSSIRKLKVHI-DADEKKHQNILEQLKESK 1804

Qy     416 HH----------------LSNFQFIWPWEEWAFVLDLPKWAPKRV-------------- 444
           ||                ||    |:  |:    :   |   :|:
Db    1805 HHADLLKDRVENLEQELILSEKNMIFQAEKSKAEIQTLKSEIQRMAQNLQDLQLELISTR 1864

Qy     445 -----FVQEILQREVRLSYWDKIKQSIEN----------------------ATALEELL 476
                ::|: : : |:|  : |  ||||                       | |:||
Db    1865 SENENLIKELKKEQERVSDLETINPSIENLLKDKEQEKVQMKEEAKITVEMLQTQLKEL- 1923

Qy     477 PPKAGPNFMYSLEEGKE--KTEEQQLSAELSRKVKEK-QTARDMIVWIEETIYPVHGFEV 533
              : ||   :|  ||:|| | :::     || |  :|:     : |     |:
Db    1924 -----NETVVSLCNDQEVSKTKEQNLGSQVQTLELEKAQLLQDLGEAKNKYII----FQS 1974

Qy     534 TLTIVVQTL------LDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQ------------- 574
           ::  : | :      |: | |    |   |:   |:  ||     :|
Db    1975 SVNALTQEVEAGKQKLEKGEKEIRTLKEQLKSQEQLVCKLAQVEGEQELWQKQKLELRNV 2034

Qy     575 VMLLSQ------------VSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQ 622
           | | |              |||    :| :  :::  :|   :   |:|  | :
Db    2035 TMALEQKVQVLQSENNTLQSTY--EALQNSHKSLESELGLIKLEKVALVERVST------ 2086

Qy     623 FHVSDQPWEI---LGNALNKTYNRISDLRKDISNITKNVLVAEKASANARV--------- 670
             :| :  |:   | : | ||     |  |: : :|: | |  :   | :
Db    2087 --ISGKEAELQRELRDMLQKTTQLSEDYNKEKNRLTEEVEVLREELQNTKAAHLKSVNQL 2144

Qy     671 --ELEAAESKLSL-------VEGE--------------------------PVLGENPA 693
             ||: |: |: |        :|||                          |: ||
Db    2145 EKELQRAQGKIKLMLKSCRQLEGEKEMLQKELSQLEAAQQQRAGSLVDSNVDEVMTEN-- 2202

Qy     694 KMKRLKSTV-EKTGEAE----------LSLRESLEAKEALLNRALSETEVLLLLLFQS-- 740
             | || |: ||  ||:          :|  |  :||| |      | ||  |   ||
Db    2203 --KALKETLEEKVKEADKYLDKYCSLLISHEELEKAKEIL------EIEVARLKSRQSRQ 2254

Qy     741 ---FLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCL 797
              :|   :| |:   || ::|              | :|  | | :
Db    2255 DLQSSPLLNSSIPGPSPNTSVSEMK----------SASGQNKASGK--------- 2290

Qy     798 STLGYLTAFTRQYASEIWPHMEKL---ESEVFS 827
                  || :| :| | ::     :| ||
Db    2291 ----------RQRSSGVWEHGKRAAPSTAETFS 2313
```

Search completed: August 20, 2002, 15:35:36
Job time: 579 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 20, 2002, 15:26:32 ; Search time 36.79 Seconds
(without alignments)
892.476 Million cell updates/sec

Title: US-09-882-986-2
Perfect score: 4374
Sequence: 1 MSNWKTLLLRIGEKGPEYGT.....EDVHPLFLQAISSALQFPLH 848

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	813	18.6	790	1	CB80_HUMAN	Q09161 homo sapien
2	279	6.4	861	1	GCR3_YEAST	P34160 saccharomyc
3	166	3.8	879	1	RA50_PYRHO	O58687 pyrococcus
4	140.5	3.2	539	1	MYS3_HYDAT	P39922 hydra atten
5	134.5	3.1	886	1	RA50_ARCFU	O29230 archaeoglob
6	134.5	3.1	1818	1	HMW2_MYCPN	P75471 mycoplasma
7	133.5	3.1	2363	1	SPCO_MOUSE	Q62261 mus musculu
8	133.5	3.1	5327	1	ACF7_MOUSE	Q9qxz0 mus musculu
9	131	3.0	1805	1	HMW2_MYCGE	P47460 mycoplasma
10	131	3.0	2663	1	CENE_HUMAN	Q02224 homo sapien
11	130.5	3.0	2364	1	SPCO_HUMAN	Q01082 homo sapien
12	130	3.0	5430	1	ACF7_HUMAN	Q9upn3 homo sapien
13	129.5	3.0	1312	1	RA50_YEAST	P12753 saccharomyc
14	129.5	3.0	1957	1	YD86_SCHPO	Q10411 schizosacch
15	128.5	2.9	1940	1	MYH3_HUMAN	P11055 homo sapien
16	128	2.9	3660	1	DMD_CHICK	P11533 gallus gall
17	128	2.9	3678	1	DMD_MOUSE	P11531 mus musculu
18	127.5	2.9	284	1	TPM1_SCHMA	P42637 schistosoma
19	127.5	2.9	1939	1	MYH1_HUMAN	P12882 homo sapien
20	127	2.9	1938	1	MYSS_CHICK	P13538 gallus gall
21	126	2.9	2230	1	GOG4_HUMAN	Q13439 homo sapien
22	126	2.9	4466	1	DYHC_ANTCR	P39057 anthocidari
23	125.5	2.9	1940	1	MYH3_RAT	P12847 rattus norv
24	125.5	2.9	2349	1	TPR_HUMAN	P12270 homo sapien
25	125.5	2.9	4466	1	DYHC_TRIGR	P23098 tripneustes
26	125	2.9	1829	1	DPOL_THEST	O33845 thermococcu
27	124.5	2.8	2470	1	TOR1_YEAST	P35169 saccharomyc
28	124	2.8	1679	1	YIO9_YEAST	P40457 saccharomyc
29	123	2.8	1044	1	YAF3_SCHPO	Q09857 schizosacch
30	123	2.8	4563	1	APB_HUMAN	P04114 homo sapien
31	122.5	2.8	1453	1	Y373_BOVIN	Q9tu23 bos taurus
32	122.5	2.8	1755	1	PEPL_MOUSE	Q9r269 mus musculu
33	122.5	2.8	1939	1	MYH4_HUMAN	Q9y623 homo sapien
34	122	2.8	2116	1	MYS2_DICDI	P08799 dictyosteli
35	121.5	2.8	1333	1	CC25_CANAL	P43069 candida alb
36	121	2.8	283	1	TPMM_LOCMI	P31816 locusta mig
37	121	2.8	1938	1	MYHD_HUMAN	Q9ukx3 homo sapien
38	121	2.8	1959	1	MYH9_CHICK	P14105 gallus gall
39	121	2.8	2469	1	TEGU_HSVSA	Q01056 herpesvirus
40	120.5	2.8	2210	1	RRPO_LYCVA	P14240 lymphocytic
41	120	2.7	499	1	C72L_ARATH	Q9ltm2 arabidopsis
42	119.5	2.7	1139	1	NGAP_HUMAN	Q9ujf2 homo sapien
43	119.5	2.7	1940	1	MYH3_CHICK	P02565 gallus gall
44	119	2.7	946	1	FTSK_CAMJE	Q46089 campylobact
45	119	2.7	1790	1	USO1_YEAST	P25386 saccharomyc

ALIGNMENTS

RESULT 1
CB80_HUMAN
ID CB80_HUMAN STANDARD; PRT; 790 AA.
AC Q09161;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 80 kDa nuclear cap binding protein (NCBP 80 kDa subunit) (CBP80).
GN NCBP1 OR NCBP OR CBP80.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 119-128; 513-522 AND 687-693.
RX MEDLINE=94349369; PubMed=8069914;
RA Izaurralde E., Lewis J., McGuigan C., Jankowska E.,
RA Darzynkiewicz E., Mattaj I.W.;
RT "A nuclear cap binding protein complex involved in pre-mRNA
RT splicing.";
RL Cell 78:657-668(1994).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=95023141; PubMed=7937105;
RA Kataoka N., Ohno M., Kangawa K., Tokoro Y., Shimura Y.;
RT "Cloning of a complementary DNA encoding an 80 kilodalton nuclear cap
RT binding protein.";
RL Nucleic Acids Res. 22:3861-3865(1994).
CC -!- FUNCTION: INVOLVED IN MEDIATING U SNRNA EXPORT FROM THE NUCLEUS.
CC BINDS TO 5'CAPPED MRNA.
CC -!- SUBUNIT: THE CAP-BINDING PROTEIN (CBC) COMPLEX IS AN HETERODIMER
CC OF CBP80 AND CBP20.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: SOME TO YEAST GCR3/STO1 PROTEIN.
CC ----------

CC ----------
DR EMBL; X80030; CAA56334.1; -.
DR EMBL; D32002; BAA06769.1; -.
DR MIM; 600469; -.
DR InterPro; IPR003890; EIF4G_cent.
DR Pfam; PF02854; MIF4G; 1.
DR SMART; SM00543; MIF4G; 1.
KW Nuclear protein; RNA-binding.
FT DOMAIN 3 20 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 790 AA; 91839 MW; F10DE7B9D16FDA0B CRC64;

Query Match 18.6%; Score 813; DB 1; Length 790;
Best Local Similarity 26.6%; Pred. No. 2.6e-42;

```
  Matches  231;  Conservative  174;  Mismatches  326;  Indels  136;  Gaps   22;

Qy       5 KTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHKIPLY 64
Db      32 ESLICKVGEK-----SACSLESNLEGLAGVLEADLPNYKSKILRLLCTVARLLPEKLTIY 86

Qy      65 GTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLI 124
Db      87 TTLVGLLNARNYNFGGEFVEAMIRQLKESLKANNYNEAVYLVRFLSDLVNCHVIAAPSMV 146

Qy     125 VVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGI 184
Db     147 AMFENFVS-----VTQEEDVP--QVRRDWYVYAFLSSLPWVGKELYEKKDAEMDRIFANT 199

Qy     185 QAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSF 244
Db     200 ESYLKRRQKTHVPMLQVW---TADKPHPQEEYLDCLWAQIQKLKKDRWQERHILRPYLAF 256

Qy     245 EAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMED 304
Db     257 DSILCEALQHNLPPF------TPP---------PHTEDSVYPMPRVIFRMFDYTDDPEGP 301

Qy     305 VQP----IDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFR--YEYLMAETLFSQILL 358
Db     302 VMPGSHSVERFVIEENLHCIIKSHWKERKTCAAQLVSYPGKNKIPLNYHIVEVIFAELFQ 361

Qy     359 LPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHL 418
Db     362 LPAPPHIDVMYTTLLIELCKLQPGSLPQVLAQATEMLYMRLDTMNTTCVDRFINWFSHHL 421

Qy     419 SNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLPP 478
Db     422 SNFQFRWSWEDWSDCLSQDPESPKPKFVREVLEKCMRLSYHQRIL----------DIVPP 471

Qy     479 ------KAGPNFMYSL-EEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYP---- 527
Db     472 TFSALCPANPTCIYKYGDESSNSLPGHSVALCLAVAFKSKATNDEIFSILKDVPNPNQDD 531

Qy     528 --VHGFE---VTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVS 582
Db     532 DDDEGFSFNPLKIEVFVQTLLHLAAKSFSHSFSALAKFHEVFKTLAESDEGKLHVLRVMF 591

Qy     583 TYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPE---NVDQFHVSDQPWEILGNALNK 639
Db     592 EVWRNHPQMIAVLVDKMIRTQIVDCAAVANWIFSSELSRDFTRLFV----WEILHSTIRK 647

Qy     640 TYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLK 699
Db     648 --------------MNKHVLKIQK-------ELEEAKEKLA------------RQHKRRS 674

Qy     700 STVEKTGEAELS-LRESLEAKEALLNRALSETEVLLLLLFQSFLGVLKERLPDPTKVRSV 758
Db     675 DDDDRSSDRKDGVLEEQIERLQEKVESAQSEQKNLFLVIFQRFIMILTEHL--------- 725

Qy     759 QDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHM 818
Db     726 -----VRCETDGTSVL-----------------TPWYKNCIERLQQIFLQHHQIIQQYM 762

Qy     819 EKLESEVFSGE-DVHPLFLQAISSALQ 844
Db     763 VTLENLLFTAELDPHILAVFQQFCALQ 789


RESULT   2
GCR3_YEAST
ID   GCR3_YEAST     STANDARD;      PRT;   861 AA.
AC   P34160;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   GCR3 protein (STO1 protein) (SUT1 protein).
GN   GCR3 OR STO1 OR SUT1 OR YMR125W OR YM8564.07 OR YM9553.01.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92380925; PubMed=1512188;
RA   Uemura H., Jigami Y.;
RT   "GCR3 encodes an acidic protein that is required for expression of
RT   glycolytic genes in Saccharomyces cerevisiae.";
RL   J. Bacteriol. 174:5526-5532(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Pandit S., Sternglanz R.;
RL   Submitted (DEC-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Guo Z., Russo P., Sherman F.;
RL   Submitted (XXX-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-492 FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Lye G., Churcher C.M., Barrell B.G., Rajandream M.A.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 342-861 FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR EXPRESSION OF GLYCOLYTIC GENES. HAS
CC       CERTAIN CHARACTERISTICS OF A TRANSCRIPTIONAL ACTIVATOR.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: SOME TO HUMAN CBP80.
CC   -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 708
CC       ONWARD AND IS SHORTER (725 AA) DUE TO A FRAMESHIFT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10224; BAA01076.1; ALT_SEQ.
DR   EMBL; L07650; -; NOT_ANNOTATED_CDS.
DR   EMBL; L27744; -; NOT_ANNOTATED_CDS.
DR   EMBL; Z49273; CAA89274.1; -.
DR   EMBL; Z48622; CAA88550.1; -.
DR   PIR; A44919; A44919.
DR   SGD; S0004732; STO1.
DR   InterPro; IPR003890; EIF4G_cent.
DR   Pfam; PF02854; MIF4G; 1.
DR   SMART; SM00543; MIF4G; 1.
KW   DNA-binding; Nuclear protein.
FT   DOMAIN       22     30       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      774    801       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      802    825       ARG/LYS-RICH (BASIC).
FT   CONFLICT    164    164       D -> V (IN REF. 3).
FT   CONFLICT    633    633       R -> I (IN REF. 3).
FT   CONFLICT    704    704       A -> R (IN REF. 3).
SQ   SEQUENCE   861 AA;  100017 MW;  EDD04907BDC9207D CRC64;


  Query Match             6.4%;  Score 279;  DB 1;  Length 861;
  Best Local Similarity   21.4%;  Pred. No. 1.5e-09;
  Matches  156;  Conservative  117;  Mismatches  242;  Indels  214;  Gaps   34;

Qy      95 DSGNCNSIRILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFY 154
Db     140 ETGPWNKIKLILRFLSIL--SPMFLVDELINIYKSLFELSIELNNLDPGNR--VPLSEAI 195

Qy     155 VICILSSLPW------GGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFF--HNGE 206
```

```
Db    196 YTNTLLNIPYLFFFNRNNDGLRTKVEE--------LLAYVEQNYLVKTTDINLLREYNGE 247

Qy    207 FESSLAE----------KDFVEDL------------------------------LDRIQ 225
          : |          |  : :|                                  :| ::
Db    248 PPYEMVELVRVVLPNVKKALINNLEQLNELFPDWNHLLTPQTGDEGFNDALTLPSVDDLK 307

Qy    226 S---LASNGWKLESV-PRPHLSFEAQL--VAGKFHELRPIKCMEQPSPPSDHSRAYSGKQ 279
          |   |  |   ::|:   |  :|   |   || |  : ||                |:|:
Db    308 SFVRLNKNFGSVDSMWKTPRYAFHVYLPNSAGNFETVVPIS------------TYAGQL 354

Qy    280 KHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCRKE--------C 331
          :| :      |  :    |: |       |  : :   ||:  |   |
Db    355 FNDII------IDLVESLEFNRKE--------VARQVITLDLFFKAGIFTEPGESIAQLI 400

Qy    332 ASYMAN-LPVTFRYEYLMAETLFSQILLLP--QPPFKTLYYTLVIMDLCKALPGAFPAVV 388
          |:|  | |  ||: | |  ||:   |  ||    ||   :|||:: |:|:  | |   |
Db    401 ATYEENPLAPTFKIEDLAIETILGLIFKLPSVSQPF-AYFYTLLV-DICQNSPKAIAPVF 458

Qy    389 AGAVRALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWA-----PKR 443
          | |  :  :  || | : | :|||  :||| | | | ||  |  |:      ||
Db    459 GRAFRFFYSHLDSLDFELKLRYLDWFSIQMSNFNFSWKWNEWED--DSIKFGKYFYNPKV 516

Qy    444 VFVQEILQREVRLSYWDKIKQSIENATALEELLPPK------------------------ 479
          | : ::|:|:||:          | : :|: || :
Db    517 NFAKNLIQKELRLT---------SNFSEVEDSLPQEFTKYLDTSYIPRDQLINYYQSLFT 567

Qy    480 --------AGPNFMYSLEEG--KEKTEEQQL----SAELSRKVKEKQTARDMIVWIEETI 525
                 | :|  :||   | |  : |      |  ||:| | ::       |   :
Db    568 GYTVEEDSVRKNDLYFRQEGVPMENTVRKILDYTHKANNSREVTELES-------ILGEL 620

Qy    526 YPVHGFEVT------LTIVVQTLLDIGSKSFTH----LVTVLERYGQVFSKLCPD-NDKQ 574
            :|  ::      : ::|| : | ||:| :|    :  : |    :|:|:  |   |:
Db    621 KNEYGSIISDFNRFVIILLVQAVTDSGSRSLSHANKYINDLKEDLKTIFAKIELDIETKE 680

Qy    575 VMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSP----------------- 617
          :::  | |:|  | |   :  |      |:::: |  ::|:
Db    681 YIIIEAVLTFWNANPQTGFLVADAFKYAGLLTSRTIFTFIFNETGLKNNGLIEATAIEAV 740

Qy    618 -ENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAE 676
           |: |  :|::     ||     : |:      |:| | ::|       ||  ::|
Db    741 FRNLSQ-QISEE--NESGNNFEFVFERLC----TIANSTIDLL-----DVNADEDIE--- 785

Qy    677 SKLSLVEGE 685
          :  | ||
Db    786 --IPKVNGE 792


RESULT   3
RA50_PYRHO
ID   RA50_PYRHO     STANDARD;      PRT;   879 AA.
AC   O58687;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA double-strand break repair rad50 ATPase.
GN   RAD50 OR PH0929.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC       rad50/mre11 complex possesses single-strand endonuclease activity
CC       and ATP-dependent double-strand-specific exonuclease activity.
CC       Rad50 provides an ATP-dependent control of mre11 by unwinding
CC       and/or repositioning DNA ends into the mre11 active site (By
CC       similarity).
CC   -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP000004; BAA30025.1; -.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   InterPro; IPR001238; RecF.
DR   InterPro; IPR003405; SMC_C.
DR   Pfam; PF00470; RecF; 1.
DR   Pfam; PF02483; SMC_C; 1.
KW   DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT   NP_BIND      30     37       ATP (BY SIMILARITY).
FT   DOMAIN      141    744       COILED COIL (POTENTIAL).
SQ   SEQUENCE   879 AA;  103673 MW;  FD4E30FD1BBCDB29 CRC64;


  Query Match             3.8%;  Score 166;  DB 1;  Length 879;
  Best Local Similarity   19.3%;  Pred. No. 0.013;
  Matches  106;  Conservative  105;  Mismatches  191;  Indels  146;  Gaps   25;

Qy    213 EKDFVEDLLDRIQSLASNGWKLESVPR-----PHLSFEAQLVAGKFHELRPIKCMEQPSP 267
          || | | :|: |::::|| :|| :  | : :  |: :  :|
Db    202 EKSFTE-VLNEIRNISSN------LPRLRRELEGIKEEVKTLEATFNSITELKL------ 248

Qy    268 PSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGC 327
                :||:      |  :|||:|      | :  : ::  | |          |
Db    249 ---RLGELNGKK-----GRLEERIRQLESGIEEKRKKSKELEEVVKE---------LPEL 291

Qy    328 RKECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAV 387
          |:   |     : |: |||: :                          :| | |      :
Db    292 EKKETEY--RRLIEFKDEYLVKKN----------------------ELEKRL-----GI 321

Qy    388 VAGAVRALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQ 447
          ::  :: :  ||   | ||:   | |    |   |                      :
Db    322 LSNRLQEVKRKIK--DAESKVARIRWIEERLKEIQ----------------------E 355

Qy    448 EILQREVRLSYWD---KIKQSIEN-ATALEELLPPKAGPNFMYSLEEGKEKTEEQ----- 498
          :|:: | |:  ::   ::|  :|:  : |  | | |     :| ||  |:: ||:
Db    356 KIMKLEPRVREFEDAMRLKAQMESLKSKLGGLEPEKINEKLLY-LENRKKELEEEIDKIT 414

Qy    499 QLSAELSRKVKEKQTARDMIVWIEET--IYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTV 556
          :   ||::: |::: |   |: ::: .   || | |:|                |   :
Db    415 RKIGELNQRSKDRRLA---IIELKKARGKCPVCGRELT---------------EEHKADL 456

Qy    557 LERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAI-VRWVF 615
          | :|    | :    :|::     :  :  :  :  : |:   : :::| | :|
Db    457 LRKYSLELSSI----EKEIQEAKALERQLRAEFRKVENELSRLSSLKTIADQIIEIRERL 512

Qy    616 SPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNIT--KNVLVAEKASANARVELE 673
          |  |::       : :|:| :  ||     :  |:|::: :   ||        |    :|::
Db    513 SKINLEDLKRDKEEYELLKSESNKLKGEVESLKKEVNELNDYKN------ESTKLEIEID 566

Qy    674 AAESKLSLVE------GEPVLGENPAKMKRL-----KSTVEKTGEAEL-SLRESLEAKEA 721
          |: :|| :|      |   : |   ::: |     |    |  | ||  : |||: :
Db    567 KAKKELSEIEDRLLRLGFKTIDELSGRIRELEKFHNKYIEAKNAEKELRDILESLKDERE 626

Qy    722 LLNRALSE 729
          |::|  |
Db    627 ELDKAFEE 634
```

```
RESULT   4
MYS3_HYDAT
ID   MYS3_HYDAT     STANDARD;      PRT;   539 AA.
AC   P39922;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, clone 203 (Fragment).
OS   Hydra attenuata (Hydra) (Hydra vulgaris).
OC   Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
OC   Hydridae; Hydra.
OX   NCBI_TaxID=6087;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nakano M.Y., Stidwill R.P.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC       ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC   -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY. STRONGEST,
CC       TO OTHER NON MUSCLE MYOSINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L35595; AAA29216.1; -.
DR   HSSP; P24733; 1WDC.
DR   InterPro; IPR002928; Myosin_tail.
DR   Pfam; PF01576; Myosin_tail; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   ATP-binding.
FT   NON_TER       1      1
FT   DOMAIN       <1     38       GLOBULAR HEAD.
FT   DOMAIN       39   >539       RODLIKE TAIL.
FT   DOMAIN       39    520       COILED COIL (POTENTIAL).
FT   NON_TER     539    539
SQ   SEQUENCE   539 AA;  62319 MW;  9C5AD56640609390 CRC64;


  Query Match             3.2%;  Score 140.5;  DB 1;  Length 539;
  Best Local Similarity   23.2%;  Pred. No. 0.23;
  Matches   96;  Conservative   60;  Mismatches  156;  Indels  101;  Gaps   19;

Qy     442 KRVFVQEI----LQREVR----LSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKE 493
           |::  | |    :|| ||    | :|   |      | :: ||      : | : ||  |
Db       5 KKMVAQRIGLAVIQRNVRKYLFLRHWSWWKL----YTKVQPLLSVARAEDEMRAKEEELE 60

Qy     494 KTEEQ-QLSAELSRKVKEK--------------QTARDMIVWIEETIYPVHGFEVTLTI 537
             :|| :  ||  :|::|:               |   | :: ||: :          :
Db      61 AAKEQLKKDAEAKKKMEEELTEAMAQKEKLYASLQAETDRLITIEDKL----------L 109

Qy     538 VVQTLLDIGSKSFT----------HLVTVL-ERYGQVFSKLCPDNDKQVMLLSQVSTYWK 586
            :||: |    |            | | || |:  :   |:    :|   | | :|   :
Db     110 NLQTVKDKLESSLNEALEKLDGEEHSVLVLEEKIQEAEEKIDELTEKTEELQSNISRL-E 168

Qy     587 NNVQMTAVAIDRMMGYRLVSNQAIVRW--VFSPENVDQFHVSD---------QPWEILGN 635
             |     || :       |: | :     |  | :: || :          |  |   |
Db     169 TEKQNRDKQIDTL-------NEDIRKQDETISKMNAEKKHVDEELKDRTEQLQAAEDKCN 221

Qy     636 ALNKTYNR----ISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGEN 691
           |||| |:    | :: :|:     : :  ||       :|:    |||  |
Db     222 NLNKTKNKLESSIREIEQDLKKEKDSKMKLEKEKKKVESDLKDNRDKLSETE-------- 273

Qy     692 PAKMKRLKSTVEKTGEAELSL------RESLEAKEALLNRALSETEVLLLLLFQSFLGVL 745
             ||| | :   : | |:      :| ||:: : | | : |    ||   :     |
Db     274 ----TRLKETQDLVTKREKSISDLENAKEGLESQISQLQRKIQE----LLAKIEE----L 321

Qy     746 KERLPDPTKVRSVQDLKSIGAEDDKPSAMD-VDSENGNPKKSCEVG-EREQWC 796
```

```
           :| | :  |:|   :|:    |       | ::: |      ||| :|| |
Db     322 EEELENERKLRQKSELQRKELESRIEELQDQLETAGGATSAQVEVGKKREAEC 374


RESULT   5
RA50_ARCFU
ID   RA50_ARCFU     STANDARD;      PRT;   886 AA.
AC   O29230;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA double-strand break repair rad50 ATPase.
GN   RAD50 OR AF1032.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC       rad50/mre11 complex possesses single-strand endonuclease activity
CC       and ATP-dependent double-strand-specific exonuclease activity.
CC       Rad50 provides an ATP-dependent control of mre11 by unwinding
CC       and/or repositioning DNA ends into the mre11 active site (By
CC       similarity).
CC   -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001032; AAB90211.1; -.
DR   TIGR; AF1032; -.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001238; RecF.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00470; RecF; 1.
KW   DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT   NP_BIND      31     38       ATP (BY SIMILARITY).
FT   DOMAIN      148    728       COILED COIL (POTENTIAL).
SQ   SEQUENCE   886 AA;  103633 MW;  D35641D499AA8B58 CRC64;


  Query Match             3.1%;  Score 134.5;  DB 1;  Length 886;
  Best Local Similarity   19.0%;  Pred. No. 1.1;
  Matches  106;  Conservative  107;  Mismatches  176;  Indels  169;  Gaps   25;

Qy     376 LCKALPGAFPAVVAGAVRALFEKISDL--DMESRTRLILWFSHHLSNFQFIWPWEEWAFV 433
           || |       | ||:    :|  :  | ||| |:|   :  : :::    |:   |
Db     114 LCPA------HVFTGAIYVRQGEIDSIIRDDESRERIIRQIT-RIEDYEN--AWKNLGAV 164

Qy     434 LDLPKWAPKRVFVQEIL--------QREVRLSYWDKIKQSIENATALEELLPPKAGPNFM 485
```

```
Db    165 IRMLEREKER--LKEFLSQEEQIKRQKEEKKAEIERISEEIKSIESLREKLSEE-----V 217

Qy    486 YSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDI 545

Db    218 RNLESRLKELEEHKSRLESLR--KQESSVLQEVRGLEEKL---RELEKQLKEVVERIEDL 272

Qy    546 --GSKSFTHLVTVLERYGQVFSKLCPDNDKQVM--------------LLSQVSTYWKNN 588

Db    273 EKKAKEVKELKPKAERY-SILEKLLSEINQALRDVEKREGDLTREAAGIQAQLKKAEEDN 331

Qy    589 VQMTAVA-------------------------IDRMMGYRLVSNQAIVRWVFSPENVDQ 622

Db    332 SKLEEITKRIEELERELERFEKSHRLLETLKPKMDRMQGIKAKLEEKNL----TPDKVE- 386

Qy    623 FHVSDQPWEILGNALNKTYNRISDLRKDISNITKNV--LVAEKASANAR--------VEL 672

Db    387 ---------------KMYDLLSKAKEEEKEITEKLKKLIAKKSSLKTRGAQLKKAVEEL 430

Qy    673 EAAE-------SKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLNR 725

Db    431 KSAERTCPVCGRELDEEHRKNIMAEYTREMKRIAEELAKADEIEKKLKERLEKVE----K 486

Qy    726 ALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKK 785

Db    487 ALEKQET-----------VLKYR-------QMVDELKAL---ENELSSHDAEKLSAESEE 525

Qy    786 SCEVGER-------EQWCLSTLGYLTAFT--------------------RQYASEIWPH 817

Db    526 YRKVKERLDGLRGQQKILLSSASRIKELKSSLREIEEALKNVESERGELHRKIREEGFES 585

Qy    818 MEKLESEVFSGEDVHPLF 835

Db    586 LEELEREV---QSLRPFY 600


RESULT   6
HMW2_MYCPN
ID   HMW2_MYCPN     STANDARD;      PRT;  1818 AA.
AC   P75471;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cytadherence high molecular weight protein 2 (Cytadherence accessory
DE   protein 2).
GN   HMW2 OR MPN310 OR MP526.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97252497; PubMed=9098066;
RA   Krause D.C., Proft T., Hedreyda C.T., Hilbert H., Plagens H.,
RA   Herrmann R.;
RT   "Transposon mutagenesis reinforces the correlation between Mycoplasma
RT   pneumoniae cytoskeletal protein HMW2 and cytadherence.";
RL   J. Bacteriol. 179:2668-2677(1997).
CC   -!- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC       STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC       CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC       PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC       IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC       SIMILARITY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000051; AAB96174.1; -.
DR   EMBL; U59896; AAB52527.1; -.
DR   HSSP; P04083; 1BO9.
KW   Cytadherence; Structural protein; Coiled coil; Complete proteome.
FT   DOMAIN       31    880       COILED COIL (POTENTIAL).
FT   DOMAIN      919   1607       COILED COIL (POTENTIAL).
FT   DOMAIN     1644   1755       COILED COIL (POTENTIAL).
FT   DOMAIN     1786   1817       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1818 AA;  215622 MW;  66DF4B08F0FCFBC0 CRC64;


  Query Match             3.1%;  Score 134.5;  DB 1;  Length 1818;
  Best Local Similarity   19.1%;  Pred. No. 3;
  Matches  158;  Conservative  104;  Mismatches  256;  Indels  309;  Gaps   34;

Qy    121 ASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERV 180

Db    235 AELQNVYETIAQNQA----------NFQKQCDAY---------WAQLKQVEQQIQTTKQE 275

Qy    181 LVGIQAYLSIRKNSSTSGLNFFHNG------EFESSLAEKDFV--EDLLDRIQSLASNGW 232

Db    276 LVDEESTLKVRLNDA----DFYINSRLAELDDLTSKINERDFVSKEQAQDVKASLANLTK 331

Qy    233 KLESVPRPHLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIR 292

Db    332 EKERLSAEKDSFE-------------------------------------------RLR 347

Qy    293 RLNIFPANKMEDVQPI-------DRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFRYE 345

Db    348 NTALNDINRMEQENALFAKHLEQQQYEFERKQQESLLKLETEHKQLQKRIG--------E 399

Qy    346 YLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEK---ISDL 402

Db    400 FKIESEAKSEALLIQE-----------------------------RELLEKRREIDDL 428

Qy    403 --------DMESRTRLILWFSH-----HLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEI 449

Db    429 LTQASLEYEQQRRTNQVLKEKHRQVQQHFQN----------LVHAKKKLDQKRHYLAE- 476

Qy    450 LQREVRLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEE----------QQ 499

Db    477 -QKRIDEEQIFKLKEKI--ATERREL-------EKLYLVKKQKQDQKENDLLIFEKQLRQ 526

Qy    500 LSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLER 559

Db    527 YQADFENEIEEKQNE--------------------LFASQKSL---QKSFTQLKNKEAE 562

Qy    560 YGQVFSKLCPD-----------NDKQVML-------------LSQVSTYWKNNVQMTAVA 595

Db    563 LNQKAQKIAEDWAHLKQNKHHHADLEIFLEGEFNHLQQEKHKLLEARTQFDNRVSLLSAR 622

Qy    596 IDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPW----------EILGNALNKTYNRI 644

Db    623 FKQKQA-ELVKQKQSLEQLTAAFNKEQEAV-ERDWKDRLANLEKQKEMLG-------DKV 673

Qy    645 SDLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEK 704

Db    674 HQFDENSLNISKKLAERELAIKFKEKELEAAQKQLSLD------NNNNA---GLKLQLDK 724

Qy    705 TGEAELSLRESLEA-KEALLN-------------------------------RALSET 730

Db    725 LSESLKTERLELEASKERILDFYDESSRRIADYESDLQARLAEVKTLEKNQQETAAKSER 784
```

```
Qy    731 EVLLLL-----LFQSFLGVLKERLPDPTKVR-----SVQDLKSIGAEDDKPS-------- 772
Db    785 ELKVALEKLNQAKKAFLQIRKQQLLEIASVKQQLAQKANLLKNQQAELDKQTEELEAAFL 844

Qy    773 AMDVDSEN-----GNPKKSCEVGEREQWCLSTLGYLTAFTRQYASEI 814
Db    845 EQDTDKKELEKALHSVKSKQELLERER------SFLLQKQREFAEHV 885


RESULT   7
SPCO_MOUSE
ID   SPCO_MOUSE     STANDARD;      PRT;  2363 AA.
AC   Q62261;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Spectrin beta chain, brain 1 (Spectrin, non-erythroid beta chain 1)
DE   (Beta-II spectrin) (Fodrin beta chain).
GN   SPTBN1 OR SPTB2 OR SPNB2 OR SPNB-2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=93240985; PubMed=8479293;
RA   Ma Y., Zimmer W.E., Riederer B.M., Goodman S.R.;
RT   "The complete amino acid sequence for brain beta spectrin (beta
RT   fodrin): relationship to globin sequences.";
RL   Brain Res. Mol. Brain Res. 18:87-89(1993).
RN   [2]
RP   STRUCTURE BY NMR OF 2199-2304.
RX   MEDLINE=94268558; PubMed=8208297;
RA   Macias M.J., Musacchio A., Ponstingl H., Nilges M., Saraste M.,
RA   Oschkinat H.;
RT   "Structure of the pleckstrin homology domain from beta-spectrin.";
RL   Nature 369:675-677(1994).
RN   [3]
RP   STRUCTURE BY NMR OF 2199-2304.
RX   MEDLINE=97342712; PubMed=9199409;
RA   Nilges M., Macias M.J., O'Donoghue S.I., Oschkinat H.;
RT   "Automated NOESY interpretation with ambiguous distance restraints:
RT   the refined NMR solution structure of the pleckstrin homology domain
RT   from beta-spectrin.";
RL   J. Mol. Biol. 269:408-422(1997).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 2199-2304.
RX   MEDLINE=96030773; PubMed=7588597;
RA   Hyvoenen M., Macias M.J., Nilges M., Oschkinat H., Saraste M.,
RA   Wilmanns M.;
RT   "Structure of the binding site for inositol phosphates in a PH
RT   domain.";
RL   EMBO J. 14:4676-4681(1995).
CC   -!- FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION,
CC       INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS
CC       THUS CANDIDATE FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE
CC       CYTOSKELETON AT THE MEMBRANE.
CC   -!- SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS
CC       ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO
CC       TETRAMERS.
CC   -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC   -!- SIMILARITY: CONTAINS 17 SPECTRIN REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M74773; AAC42040.1; -.
DR   PDB; 1BTN; 08-MAR-96.
DR   PDB; 1MPH; 16-JUN-97.
DR   MGD; MGI:98388; Spnb2.
DR   InterPro; IPR001589; Actinin_act_bind.
DR   InterPro; IPR001715; Calponin_hom.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR001605; Spectrin_PH.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00435; spectrin; 18.
DR   PRINTS; PR00683; SPECTRINPH.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00150; SPEC; 16.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00020; ACTININ_2; 1.
DR   PROSITE; PS50021; CH; 2.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
KW   Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein;
KW   3D-structure.
FT   DOMAIN        1    275       ACTIN-BINDING (BY SIMILARITY).
FT   DOMAIN       54    158       CH 1.
FT   DOMAIN      173    275       CH 2.
FT   REPEAT      276    384       SPECTRIN 1.
FT   REPEAT      385    498       SPECTRIN 2.
FT   REPEAT      499    608       SPECTRIN 3.
FT   REPEAT      609    714       SPECTRIN 4.
FT   REPEAT      715    819       SPECTRIN 5.
FT   REPEAT      820    925       SPECTRIN 6.
FT   REPEAT      926   1032       SPECTRIN 7.
FT   REPEAT     1033   1139       SPECTRIN 8.
FT   REPEAT     1140   1245       SPECTRIN 9.
FT   REPEAT     1246   1350       SPECTRIN 10.
FT   REPEAT     1351   1462       SPECTRIN 11.
FT   REPEAT     1463   1562       SPECTRIN 12.
FT   REPEAT     1563   1668       SPECTRIN 13.
FT   REPEAT     1669   1775       SPECTRIN 14.
FT   REPEAT     1776   1881       SPECTRIN 15.
FT   REPEAT     1882   1987       SPECTRIN 16.
FT   REPEAT     1988   2132       SPECTRIN 17.
FT   DOMAIN     2196   2306       PH.
SQ   SEQUENCE   2363 AA;  274420 MW;  64C9E4BD26BBC7B8 CRC64;


  Query Match             3.1%;  Score 133.5;  DB 1;  Length 2363;
  Best Local Similarity   17.3%;  Pred. No. 5;
  Matches  174;  Conservative  152;  Mismatches  284;  Indels  395;  Gaps   46;

Qy    118 IQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEI 177
Db    298 IETEKMIEKYESLASDLLEWIEQ-------------TIIILNNRKFANS---------- 333

Qy    178 ERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQS-LASNGWKLES 236
Db    334 ---LVGVQQQL--------QAFNTYRTVEKPPKFTEKGNLEVLLFTIQSKMRANNQKV-Y 381

Qy    237 VPRPHLSFEAQLVA-----------------GKFHELRPIKCMEQPSPPSDHSRAY--- 275
Db    382 MPR-----EGKLISDINKAWERLEKAEHERELALRNELIRQEKLEQLARRFDRKAAMRET 436

Qy    276 ----------------------SGKQKHDA----LTRYPQRIRRLNIFPANKME-----D 304
Db    437 WLSENQRLVSQDNFGFDLPAVEAATKKHEAIETDIAAYEERVQAV-VAVARELEAENYHD 495

Qy    305 VQPI----DRFV-VEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLF------ 353
Db    496 IKRITARKDNVIRLWEYLLELL----RARRQ------RLEMNLGLQKIFQEMLYIMDWMD 545

Qy    354 -SQILLLPQPPFKTLYYTLVIMDLCK--ALPGAFPAVVAGAVRAL-------------- 395
```

```
Db    546 EMKVLLLSQDYGKHL---LGVEDLLQKHALVEADIAIQAERVRGVNASAQKFATDGEGYK 602

Qy    396 ----------------FEKISDLDMESRTRL----ILWFSHHLSNFQFIWPWEE---WA 431
                          ::::  |  | | ||     ||        :| |   |   |
Db    603 PCDPQVIRDRVAHMEFCYQELCQLAAERRARLEESRRLW--------KFFWEMAEEEGW- 653

Qy    432 FVLDLPKWAPKRVFVQEILQREVRLSYWD------KIKQSIENATALEELLPPKAGPNFM 485
                          | ::|  ||  |       : : :    | |: :  ::| :|
Db    654 ----------------IREKEKILSSDDYGKDLTSVMRLLSKHRAFEDEMSGRSG-HFE 695

Qy    486 YSLEEGKEKTEEQQLSAELSRKVKEK------------QTARDMIVWIEETIYPVHGFEV 533
          :::||::   |:   :|   |::|:             | :   |  :||    :| |:
Db    696 QAIKEGEDMIAEEHFGSE---KIRERIIYIREQWANLEQLSAIRIKRLEEASL-LHQFQA 751

Qy    534 TLTIVVQTLLDI----GSKSFTH----LVTVLERYGQVFSKL----------------- 567
            :  :|||     |    |      ::::::  |  ::
Db    752 DADDIDAWMLDILKIVSSNDVGHDEYSTQSLVKKHKDVAEEITNYRPTIDTLHEQASALP 811

Qy    568 ------------------CPDNDKQVMLLSQV----------------------STYW 585
                           | :  :   |  |                        |
Db    812 QAHAESPDVKGRLAGIEERCKEMAELTRLRKQALRDTLALYKMFSEADACELWIDEKEQW 871

Qy    586 KNNVQM------------------------TAVAIDRMMGYRLVSN------------- 607
          ||:|:                        : ||:   :  :|: |
Db    872 LNNMQIPEKLEDLEVIQHRFESLEPEMNNQASRVAVVNQIARQLMHNGHPSEKEIRAQQD 931

Qy    608 QAIVRW-------------VFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRK----- 649
          :   ||              : |  ::  :|:            |:| : | :  |
Db    932 KLNTRWSQFRELVDRKKDALLSALSIQNYHLE----------CNETKSWIREKTKVIEST 981

Qy    650 -DISNITKNVLVAEKASANARVELEAAESKLSLV--EGEPVLGENPAKMKRLKSTVEKTG 706
           |: |    |:  ::       :| | |:||| :  | | :  |:| : : : | : :
Db    982 QDLGNDLAGVMALQRKLTGMERDLVAIEAKLSDLQKEAEKLESEHPDQAQAILSRLAEIS 1041

Qy    707 EAELSLRESLEAKEALLNRALSETEVLLLLL--FQSFLGVLKERLPDPTKVRSVQDLKSI 764
          :    :: :|: :|| |  | |: :  |  |  |||:|   :  :
Db   1042 DVWEEMKTTLKNREASLGEA-SKLQQFLRDLDDFQSWLSRTQTAI-------------- 1085

Qy    765 GAEDDKPSAM--------------DVDSENGNPKKSCEVGEREQWCLSTLGYLTA---F 806
           | :| |: :                ::|:   : :|  ::||     : | |   |   |
Db   1086 -ASEDMPNTLTEAEKLLTQHENIKNEIDNYEEDYQKMRDMGE-----MVTQGQTDAQYMF 1139

Qy    807 TRQYASEI---WPHMEKL---ESEVFSGEDVHPLFLQAISSALQF 845
           ||    :   |  : |:       : |    :  ||:    |  |
Db   1140 LRQRLQALDTGWNELHKMWENRQNLLSQSHAYQQFLRDTKQAEAF 1184


RESULT   8
ACF7_MOUSE
ID   ACF7_MOUSE     STANDARD;      PRT;  5327 AA.
AC   Q9QXZ0; P97394; P97395; P97396;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Actin cross-linking family protein 7 (Microtubule actin crosslinking
DE   factor) (MACF).
GN   ACF7 OR ACLP7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   STRAIN=BALB/C;
RX   MEDLINE=20069791; PubMed=10601340;
RA   Leung C.L., Sun D., Zheng M., Knowles D.R., Liem R.K.H.;
RT   "Microtubule actin cross-linking factor (MACF): a hybrid of dystonin
RT   and dystrophin that can interact with the actin and microtubule
RT   cytoskeletons.";
RL   J. Cell Biol. 147:1275-1286(1999).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC   STRAIN=BALB/C; TISSUE=Brain;
RX   MEDLINE=97124842; PubMed=8954775;
RA   Bernier G., Mathieu M., De Repentigny Y., Vidal S.M., Kothary R.;
RT   "Cloning and characterization of mouse ACF7, a novel member of the
RT   dystonin subfamily of actin binding proteins.";
RL   Genomics 38:19-29(1996).
CC   -!- FUNCTION: F-ACTIN-BINDING PROTEIN WHICH MAY PLAY A ROLE IN CROSS-
CC       LINKING ACTIN TO OTHER CYTOSKELETAL PROTEINS. ALSO BINDS TO
CC       MICROTUBULES.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1, 2 (SHOWN HERE) AND 3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED MAINLY IN LUNG, BRAIN, SPINAL CORD,
CC       SKELETAL AND CARDIAC MUSCLE, AND SKIN.
CC   -!- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 37 SPECTRIN REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF150755; AAD32244.1; -.
DR   EMBL; U67203; AAC52988.1; -.
DR   EMBL; U67204; AAC52989.1; -.
DR   EMBL; U67205; AAC52990.1; -.
DR   HSSP; Q01082; 1BKR.
DR   MGD; MGI:108559; Aclp7.
DR   InterPro; IPR001589; Actinin_act_bind.
DR   InterPro; IPR001715; Calponin_hom.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR003108; GAS2.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00036; efhand; 2.
DR   Pfam; PF02187; GAS2; 1.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00435; spectrin; 37.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00054; EFh; 2.
DR   SMART; SM00243; GAS2; 1.
DR   SMART; SM00326; SH3; 1.
DR   SMART; SM00150; SPEC; 32.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR   PROSITE; PS50021; CH; 2.
DR   PROSITE; PS50002; SH3; FALSE_NEG.
DR   PROSITE; PS00018; EF_HAND; 2.
KW   Actin-binding; Cytoskeleton; Calcium-binding; Repeat; SH3 domain;
KW   Alternative splicing.
FT   DOMAIN        1    295       ACTIN-BINDING (BY SIMILARITY).
FT   DOMAIN       78    181       CH 1.
FT   DOMAIN      194    295       CH 2.
FT   REPEAT      314    355       SPECTRIN 1.
FT   REPEAT      591    623       SPECTRIN 2.
FT   REPEAT      680    784       SPECTRIN 3.
FT   REPEAT      786    800       SPECTRIN 4.
FT   DOMAIN      871    923       SH3.
FT   REPEAT     1250   1272       SPECTRIN 5.
FT   REPEAT     1287   1342       SPECTRIN 6.
FT   REPEAT     1458   1534       SPECTRIN 7.
FT   REPEAT     1593   1660       SPECTRIN 8.
FT   REPEAT     1817   1886       SPECTRIN 9.
FT   REPEAT     1934   2044       SPECTRIN 10.
FT   REPEAT     2262   2282       SPECTRIN 11.
FT   REPEAT     2376   2397       SPECTRIN 12.
FT   REPEAT     2400   2509       SPECTRIN 13.
```

```
FT   REPEAT      2557    2617        SPECTRIN 14.
FT   REPEAT      2620    2727        SPECTRIN 15.
FT   REPEAT      2730    2837        SPECTRIN 16.
FT   REPEAT      2840    2944        SPECTRIN 17.
FT   REPEAT      2946    2972        SPECTRIN 18.
FT   REPEAT      2986    3029        SPECTRIN 19.
FT   REPEAT      3086    3162        SPECTRIN 20.
FT   REPEAT      3195    3273        SPECTRIN 21.
FT   REPEAT      3276    3382        SPECTRIN 22.
FT   REPEAT      3385    3491        SPECTRIN 23.
FT   REPEAT      3494    3518        SPECTRIN 24.
FT   REPEAT      3603    3709        SPECTRIN 25.
FT   REPEAT      3722    3817        SPECTRIN 26.
FT   REPEAT      3869    3930        SPECTRIN 27.
FT   REPEAT      3933    4039        SPECTRIN 28.
FT   REPEAT      4042    4149        SPECTRIN 29.
FT   REPEAT      4152    4258        SPECTRIN 30.
FT   REPEAT      4261    4368        SPECTRIN 31.
FT   REPEAT      4371    4477        SPECTRIN 32.
FT   REPEAT      4480    4587        SPECTRIN 33.
FT   REPEAT      4594    4695        SPECTRIN 34.
FT   REPEAT      4698    4804        SPECTRIN 35.
FT   REPEAT      4807    4872        SPECTRIN 36.
FT   REPEAT      4917    4941        SPECTRIN 37.
FT   CA_BIND     4987    4998        EF-HAND 1 (POTENTIAL).
FT   CA_BIND     5023    5034        EF-HAND 2 (POTENTIAL).
FT   DOMAIN      5173    5180        POLY-SER.
FT   DOMAIN      5252    5267        4 X 4 AA TANDEM REPEATS OF [GS]-S-R-[AR].
FT   VARSPLIC       1      73        MSSSDEETLSERSCRSERSCRSERSYRSERSGSLSPCPPGD
FT                                   TLPWNLPLHEQKKRKSQDSVLDPAERAVVRVA -> EKEFV
FT                                   QAYEDVLERYK (IN ISOFORM 1).
FT   VARSPLIC       1     181        MSSSDEETLSERSCRSERSCRSERSYRSERSGSLSPCPPGD
FT                                   TLPWNLPLHEQKKRKSQDSVLDPAERAVVRVADERDRVQKK
FT                                   TFTKWVNKHLMKVRKHINDLYEDLRDGHNLISLLEVLSGIK
FT                                   LPREKGRMRFHRLQNVQIALDFLKQRQVKLVNIRNDDITDG
FT                                   NPKLTLGLIWTIILHFQ -> MGNSLGCVKEPKESIAVPEK
FT                                   APISPKKRVRFKRKWRGKKILTPEASHREEALEGTGVIEET
FT                                   ETLTKLTARLPKEPGVGGAEHPPSDIFLPGDSAPNSGVGDQ
FT                                   GMIVQVKESFQAEIQTAHLLLENESSVVGGAWDSLEEGMTV
FT                                   IAHLLDNPAERNCEKSVSQLVEFPRTASCSSRAVLLPLQGE
FT                                   TAVEQGGTLLRHRHRSSTLPRTDYPSETVDQDQPSEGWSVG
FT                                   GRTKSVPSAPPTGSWIAKCSVASSIPKQSGDPIHTEPTHVG
FT                                   LVSCKGPIMPASQSDLSVSGITVSILPSSSGYGSDGLRLHG
FT                                   IRPEDTEPEKTSTPFSEEDGTLSLE (IN ISOFORM 3).
FT   CONFLICT    393     393        L -> P (IN REF. 2; AAC52990).
FT   CONFLICT   1882    1882        Q -> H (IN REF. 2).
SQ   SEQUENCE   5327 AA;  607972 MW;  5DE4FF5A6514BFFA CRC64;
```

```
  Query Match             3.1%;  Score 133.5;  DB 1;  Length 5327;
  Best Local Similarity   19.3%;  Pred. No. 16;
  Matches  194;  Conservative  141;  Mismatches  339;  Indels  329;  Gaps   48;

Qy      55 EQLPHKIPLYGTLIG------LLNLENEDFVQKLVESVHA-----NFQVALDSGNCNSIR 103
           |||         :|:|       |   | | | |   || |        :: :    | :
Db     836 EQLIQSKSSVASLVGRSKTIVQLKPRNPDHVLKSTLSVKAICDYRQIEITI----CKNDE 891

Qy     104 ILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLP 163
             :|   :     |||  |                      || :  |      :| |   |
Db     892 CVLEDNSQRTKWKVISPT-------------------GNEAMVPS-----VCFLIPPP 925

Qy     164 -WGGSELAEQVPDEIERVL-------VGIQAYLS--IRKNSST------SGLNFFHNGE 206
            |  |:| :|    ::|:       :  :: :|   :||:  |        |     ||
Db     926 NKGAIEMASRVEQSYQKVMALWHQLHINTKSLISWNYLRKDLDTVQTWSLEKLRSLAPGE 985

Qy     207 FESSLAE-----KDFVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELRPIKC 261
                :     :||::|         |:   | ||    |  | :: | | |    :|
Db     986 CHQVMKNLQAHYEDFLQD---------SHDSALFSV-ADRLRIEEEVEACKAHFQHLMKS 1035

Qy     262 MEQPSPPSDHSRAYSGKQKHD--ALTRYPQRIRRLNIFPANKMED---VQPID-RFVVEE 315
           :|        :: |  : |:    |    ||: :    ||:   |    | |  |   :|
Db    1036 LENEDKEETLAKVYISELKNIRLLLEECEQRLLKQIQSPASSKTDRDARQDITLRIAEQE 1095
```

```
Qy     316 YLLDVLFY----LNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLPQP-----PFKT 366
           :  : | :    |:    :|  ::   |         | || |:: |: :         |
Db    1096 HTQEDLQHLRSDLDAISMKCNVFLQQSP-----SGSSATTLRSELNLMVEKMDHVYGLST 1150

Qy     367 LYYT-LVIMD-LCKALPGAFPAVVAGAVRALFEKISDLDMES----RTRLILWFSH-HLS 419
           :|   |  :| : :::  |   |    ::   |:    |: :    || |  | |
Db    1151 VYLNKLKTIDVIVRSMQDAELLVKGYEIKLSQEEAVPADLSALESHRTTLQHWLSDVKDK 1210

Qy     420 NFQFIWPWEEWAFV--------------LDLPKWAPKRVFVQE-----ILQREVRLSYW 459
           |  |    ||                 |||  ::  |   :||     | | | | |
Db    1211 NSVFSVLDEEITKAKKVAEQLRHPASEPNLDLERYQEKGSQLQERWHRVIAQLETRQSEV 1270

Qy     460 DKIKQS-------------IENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELS 505
           : |::               ||  || :|:: |          |: :  :|:    ||
Db    1271 ESIQEVLRDYRACHGTLIKWIEETTAQQEMMKPG------QAEDSRVLSEQLSQQTELF 1323

Qy     506 RKVKEKQTARDMIVWIEE---TIYPVHGFEVTL----------------------TIV 538
           :::  || |      :   ||  |  :|: |                        :
Db    1324 AEIERNQTKLDQCQKFSQQYSTI--VKDYELQLMTYKAFVESQQKSPGKRRRMISSSDAI 1381

Qy     539 VQTLLDIGSKSFTHLVTVLERYGQVFS---KLCPDNDKQVM-----------LLSQVST 583
           |  :|: :: :| |||:  :: :  |   :   : :| |              ||  |||
Db    1382 TQEFMDLRTR-YTALVTLTTQHVKYISDALRRLEEEEKVVEEEKQEHVEKVKDLLGWVST 1440

Qy     584 YWKNNVQMT------------AVAIDRMMGYRLVSNQAIVRWVFSPENV---------- 620
           :|    |              |:   :::   | : :  |        :
Db    1441 LARNTQGTTTSSHTSASADIEKAILEQQVLAEELTTKKEQVSEAIKTSQIFLAKHGHKLS 1500

Qy     621 --DQFHVSDQPWEILGNALNKTY--------NRISDLRKDISNITKN------------ 657
             ::  :|:|        |||||        |::  |: ::: |:
Db    1501 EGEKEQISEQL-----RVLNKTYHDLCDGSANQLQQLQSELAQQTEQKTLQKQQDTCHKK 1555

Qy     658 -------VLVAEKA-------SANARVELEAAESKLSL--VEG---------------E 685
                 |  ||:|       |:: :  |  || |: |  ::|               |
Db    1556 LEDLCNWVGQAERALERHQGGSSSGKSSLLCAEPKVDLKDLQGDIQSHSTSFATAVKDIE 1615

Qy     686 PVLGENPAKMKRLKSTVEKTGEAELSLRESL-EAKEALLNRALSETEVLLLLLFQSFLGV 744
           | ||  |:   : |          :||| | :|||                        |
Db    1616 GFLEENQTKLSPQELT---------ALREKLHQAKEQ--------------------YEV 1646

Qy     745 LKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLT 804
           |:||    |:|   :       |:   ||:  ::|    | : |: | ::   : |
Db    1647 LQER----TRVAQKE------LEEAVTSALQQETEKS--KAATELAENKRKIDALL---- 1690

Qy     805 AFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAISSALQFPL 847
             :   :| |:          : | |  |  ||:|:  :
Db    1691 ----DWVGLLWGHL----------RESHRLAFQEWSSSLELAM 1719
```

```
RESULT   9
HMW2_MYCGE
ID   HMW2_MYCGE     STANDARD;      PRT;  1805 AA.
AC   P47460;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cytadherence high molecular weight protein 2 (Cytadherence accessory
DE   protein 2).
GN   HMW2 OR MG218.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
```

```
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
RN   [2]
RP   SEQUENCE OF 557-659 FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=94075230; PubMed=8253680;
RA   Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT   "A survey of the Mycoplasma genitalium genome by using random
RT   sequencing.";
RL   J. Bacteriol. 175:7918-7930(1993).
CC   -!- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC       STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC       CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC       PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC       IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC       SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39701; AAC71437.1; -.
DR   EMBL; U02165; AAD12447.1; -.
DR   TIGR; MG218; -.
DR   InterPro; IPR003364; Seryl_tRNA_N.
DR   Pfam; PF02403; Seryl_tRNA_N; 1.
KW   Cytadherence; Structural protein; Coiled coil; Complete proteome.
FT   DOMAIN       28    838       COILED COIL (POTENTIAL).
FT   DOMAIN      914   1591       COILED COIL (POTENTIAL).
FT   DOMAIN     1632   1723       COILED COIL (POTENTIAL).
FT   DOMAIN     1777   1804       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1805 AA;  216252 MW;  11D093AF173284FD CRC64;


  Query Match             3.0%;  Score 131;  DB 1;  Length 1805;
  Best Local Similarity   20.1%;  Pred. No. 4.9;
  Matches  176;  Conservative  139;  Mismatches  302;  Indels  258;  Gaps   41;

Qy      49 FLLQCAEQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRF 108
           :||   :|| :::        ||::| |:  ::|| |     |   :: : | |:   ::
Db     192 YLLNVIDQLYNELDQLENQKRLLSIEYENTYRELV-$ADNELQNVYENIDQNQIQFKHQY 250

Qy     109 MT-----SLLCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLP 163
           |     | |  |:       :|  |   |:    :|:          ||||:   |: |
Db     251 QTYRDELSQLERKIQLTKQELVDKE---SALRVKIDD----------ADFYINARLAELD 297

Qy     164 WGGSELA------EQVPDEIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFV 217
               :|:      :|    :|  || :        |     ||      |  :| :   :
Db     298 DVAKQLSFQDGITKQNAQHVEDKLVAL--------NKEKDRLNTQKEAFF--NLRQSALI 347

Qy     218 EDLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSG 277
           :  ::::|                  | :| |                  :| :
Db     348 D--INKLQQ-----------------ENELFAKHL---------------EHQQNEFE 371

Qy     278 KQKHDAL----TRYPQRIRRLNIF---PANKMEDVQPIDRFVVEEYLLDVLFYLNGCRKE 330
           ::: |:|    | |     ::| |    | | |::   :| : |:            |:|
Db     372 QKQSDSLLKLETEYKALQHKINEFKNESATKSEELLNQERELFEK-----------RRE 419

Qy     331 CASYMANLPVTFRYEYLMAETLFSQILLLPQ----PPFKTLYYTLVIMDLCKALPGAFPA 386
             : :     :  ||:   :   ||:|   |       |: | |    :|  : |
Db     420 IDTLLTQ--ASLEYEH---QRESSQLLKDKQNEVKQHFQNLEYAKKELDKERNLLDQQKK 474

Qy     387 VVAGAVRALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFV 446
           | : |:  | ||::    : |  |                ||   |        |:
Db     475 VDSEAIFQLKEKVA----QERKEL----------------EELYLV--------KK--- 502

Qy     447 QEILQREVRLSYWDK-IKQ---SIEN--ATALEELLPPKAGPNFMYSLEEGKEK---TEE 497
           |:  |:|  | :::| :||     ||      :||   |      :   | |||   |:
Db     503 QKQDQKENELLFFEKQLKQHQADFENELEAKQQELFEAKHALERSFIKLEDKEKDLNTKA 562

Qy     498 QQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVL 557
           ||:: | |:   :|  : |              ||: |    : |     | |    |
Db     563 QQIANEFSQLKTDKSKSAD-------------FELMLQNEYENLQQEKQKLFQER-TYF 607

Qy     558 ERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMM----GYRLVSNQAIVRW 613
           ||   | |       |:  || | |           :|::     ||:: :
Db     608 ERNAAVLSNRL--QQKREELLQQKET-----------LDQLTKSFEQERLINQREHKEL 653

Qy     614 VFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELE 673
           | | |        :  ||||       :: |  :   | :||:   | |      |:|
Db     654 VASVE---------KQKEILG-------KKLQDFSQTSLNASKNLAEREMAIKFKEKEIE 697

Qy     674 AAESKL-----------SLVEGEPVLGENPAKMKRLK--------STVEKTGEAELSLR 713
           | | :|             | :    | :  ::::  |          :::|  | ||||:
Db     698 ATEKQLLNDVNNAEVIQADLAQLNQSLNQERSELQNAKQRIADFHNDSLKKLNEYELSLQ 757

Qy     714 ESLEAKEAL---------LNRALSETEV-LLLLLFQSFLGV----------LKERLPDPT 753
           : |:  : |          |:|  | |:  |    |:|| :           :|:|| |
Db     758 KRLQELQTLEANQKQHSYQNQAYFEGELDKLNREKQAFLNLRKKQTMEVDAIKQRLSDKH 817

Qy     754 KVRSVQ----DLKS-------IGAEDDKPSAMD----------VDSENGNPKKSCEVGE 791
           :  ::|    | |:       :  : |: |  |           :| |     :|  : :
Db     818 QALNMQQAELDRKTHELNNAFLNHDADQKSLQDQLATVKETQKLIDLE-----RSALLEK 872

Qy     792 REQWCLSTLGYLTAFTRQYA--SEIWPHMEKLESE 824
           : ::  :  |:   :: : :  :|:   :| |||
Db     873 QREFAENVAGFKRHWSNKTSQLQKIYELTKKQESE 907


RESULT  10
CENE_HUMAN
ID   CENE_HUMAN     STANDARD;      PRT;  2663 AA.
AC   Q02224;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Centromeric protein E (CENP-E protein).
GN   CENPE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93024922; PubMed=1406971;
RA   Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT   "CENP-E is a putative kinetochore motor that accumulates just before
RT   mitosis.";
RL   Nature 359:536-539(1992).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=95196755; PubMed=7889940;
RA   Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT   "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT   microtubule motor.";
RL   EMBO J. 14:918-926(1995).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=98437347; PubMed=9763420;
RA   Chan G.K.T., Schaar B.T., Yen T.J.;
RT   "Characterization of the kinetochore binding domain of CENP-E reveals
RT   interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL   J. Cell Biol. 143:49-63(1998).
CC   -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC       KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC       OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC       AND/OR SPINDLE ELONGATION.
CC   -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC   -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC       CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC       QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z15005; CAA78727.1; -.
DR   PIR; S28261; S28261.
DR   HSSP; P17119; 3KAR.
DR   MIM; 117143; -.
DR   InterPro; IPR001752; kinesin.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW   Cell cycle; Centromere.
FT   DOMAIN       1    335       KINESIN-MOTOR.
FT   DOMAIN     336   2471       COILED COIL (POTENTIAL).
FT   DOMAIN    2472   2663       GLOBULAR (POTENTIAL).
FT   NP_BIND     86     93       ATP (BY SIMILARITY).
SQ   SEQUENCE   2663 AA;  312087 MW;  CEFC13880C8C8CB8 CRC64;


  Query Match             3.0%;  Score 131;  DB 1;  Length 2663;
  Best Local Similarity   17.3%;  Pred. No. 8.5;
  Matches  195;  Conservative  172;  Mismatches  373;  Indels  384;  Gaps   43;

Qy      17 EYGTSSDYKDHIETCFGVI---RREIERSGDQVLPFLLQCAEQLPHKIPLYGTLIGLLNL 73
           :|  :: | : : |   ::   |:||     |:   |:   |   |  |          |
Db     327 KYMKNTPYVNEVSTDEALLKRYRKEIMDLKKQLEEVSLETRAQAMEKDQLAQL------L 380

Qy      74 ENEDFVQK------------LVESVHANFQVALDS----------GNCNSIR-------- 103
           | :| :||            || |     |  | :          |  | ::
Db     381 EEKDLLQKVQNEKIENLTRMLVTSSSLTLQQELKAKRKRRVTWCLGKINKMKNSNYADQF 440

Qy     104 --------------ILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQ 148
                        |||  :   :||:             : |:   |: | : ||: :
Db     441 NIPTNITTKTHKLSINLLREIDESVCSE----------SDVFSNTLDTLSEIEWNPATK 489

Qy     149 PQADFYVICILSSLPWGGSELA---EQVPDEIERVLVGIQAYLSI-------RKNSSTSG 198
             :   |:||     |    ||:  | | : : ::   :          ||
Db     490 LLNQENIESELNSLRADYDNLVLDYEQLRTEKEEMELKLKEKNDLDEFEALERKTKKDQE 549

Qy     199 LNFFH--------------NGEFESSLAEKDFV----EDLLDRIQSLASN--------- 230
           :  |              | : |: |: |  :    || : ::|    :
Db     550 MQLIHEISNLKNLVKHREVYNQDLENELSSKVELLREKEDQIKKLQEYIDSQKLENIKMD 609

Qy     231 -GWKLESVPRP----HLSFEAQLVA------GKFHELRPIKCMEQPSPPSDHSRAYSGKQ 279
            : |||:  |       |:|: ||        |     ::  |:         | : ||
Db     610 LSYSLESIEDPKQMKQTLFDAETVALDAKRESAFLRSENLELKEKMK-----ELATTYKQ 664

Qy     280 KHDALTRYPQRIRRLNIFPANKMEDVQP----------IDRFVVEEYLLDVLFYLNGCR 328
             : :  |  ::        :  :::|           ||  | :: |         |
Db     665 MENDIQLYQSQLEAKKKMQVDLEKELQSAFNEITKLTSLIDGKVPKDLL---------CN 715

Qy     329 KECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPG------ 382
            |    : :|      |    | |  :::|| :                |:||
Db     716 LELEGKITDLQKELNKEVEENEALREEVILLSE--------------LKSLPSEVERLR 760

Qy     383 ------------------AFPAVV--AGAVRALFEKIS--------------------- 400
                            |  ||     |: | |:|
Db     761 KEIQDKSEELHIITSEKDKLFSEVVHKESRVQGLLEEIGKTKDDLATTQSNYKSTDQEFQ 820

Qy     401 -----DLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPK----------R 443
                :| | : :::|  :  ::             :::| | | |
Db     821 NFKTLHMDFEQKYKMVLEENERMNQ----------EIVNLSKEAQKFDSSLGALKTELS 869

Qy     444 VFVQEILQ--REV--RLSYWDKIKQSIEN----------ATALEELLPPKAGPNFMYSL 488
```

```
           ||: :  |||  ||:  :::|: :||           | : | |         :|
Db     870 YKTQELQEKTREVQERLNEMEQLKEQLENRDSPLQTVEREKTLITEKL--------QQTL 921

Qy     489 EEGKEKTEEQ--------------QLSAELSRKV-------KEKQTARDMIVWIEETIY 526
           || |  |:|:              || :::  |       :: : | : :   :|||
Db     922 EEVKTLTQEKDDLKQLQESLQIERDQLKSDIHDTVNMNIDTQEQLRNALESLKQHQETIN 981

Qy     527 P------------VHGFEVT---------LTIVVQTLLDIGSKSFTHLVT------VLER 559
                       :|  | |           : :    |: :|:   |        ::|:
Db     982 TLKSKISEEVSRNLHMEENTGETKDEFQQKMVGIDKKQDLEAKNTQTLTADVKDNEIIEQ 1041

Qy     560 YGQVFSKLCPDNDKQVMLLS------QVSTYWKNNVQMTAVAID--RMMGYRLVSNQAIV 611
             ::|| :   |: | || |      |: |  | |::||    :  |::|  |   | ||
Db    1042 QRKIFSLIQEKNELQQMLESVIAEKEQLKTDLKENIEMTIENQEELRLLGDELKKQQEIV 1101

Qy     612 RWVFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVE 671
                 :: |   :  |     |::| :|:::: : :   :: :   ::   | : |
Db    1102 A-------QEKNHAIKKEGE-----LSRTCDRLAEVEEKLKEKSQQLQEKQQQLLNVQEE 1149

Qy     672 LEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELS-------------------- 711
           :   : |:: :|      :|  | | |     :|   ||:
Db    1150 MSEMQKKINEIENL----KNELKNKELTLEHMETERLELAQKLNENYEEVKSITKERKVL 1205

Qy     712 --LRESLEAKEALLNRALSETEV-------------LLLLLFQSFLGVLKERLPDPT-KV 755
             |::| | :   |   : | |              : |  | :  |:  : : | ::
Db    1206 KELQKSFETERDHLRGYIREIEATGLQTKEELKIAHIHLKEHQETIDELRRSVSEKTAQI 1265

Qy     756 RSVQDLKS--IGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLTAFTRQYASE 813
           : |||:      ::: |   :       | ||  |  |       |:  |  | |  ::
Db    1266 INTQDLEKSHTKLQEEIPVLHEEQELLPNVKKVSETQE-------TMNELELLTEQSTTK 1318

Qy     814 IWPHMEKLESEVF--------SGEDVHPLF-----LQAISSALQ 844
             : ::| |           | |::  |      |: |  ||:
Db    1319 DSTTLARIEMERLRLNEKFQESQEEIKSLTKERDNLKTIKEALE 1362


RESULT  11
SPCO_HUMAN
ID   SPCO_HUMAN     STANDARD;      PRT;  2364 AA.
AC   Q01082; Q16057; O60837;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Spectrin beta chain, brain 1 (Spectrin, non-erythroid beta chain 1)
DE   (Beta-II spectrin) (Fodrin beta chain).
GN   SPTBN1 OR SPTB2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG ISOFORM).
RC   TISSUE=Brain;
RX   MEDLINE=92406787; PubMed=1527002;
RA   Hu R.J., Watanabe M., Bennett V.;
RT   "Characterization of human brain cDNA encoding the general isoform of
RT   beta-spectrin.";
RL   J. Biol. Chem. 267:18715-18722(1992).
RN   [2]
RP   SEQUENCE OF 293-1544 FROM N.A.
RX   MEDLINE=94010920; PubMed=8406479;
RA   Chang J.G., Scarpa A., Eddy R.L., Byers M.G., Harris A.S.,
RA   Morrow J.S., Watkins P., Shows T.B., Forget B.G.;
RT   "Cloning of a portion of the chromosomal gene and cDNA for human beta-
RT   fodrin, the nonerythroid form of beta-spectrin.";
RL   Genomics 17:287-293(1993).
RN   [3]
RP   SEQUENCE OF 2087-2168 FROM N.A. (SHORT ISOFORM).
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=20267884; PubMed=10806113;
RA   Hayes N.V.L., Scott C., Heerkens E., Ohanian V., Maggs A.M.,
RA   Pinder J.C., Kordeli E., Baines A.J.;
RT   "Identification of a novel C-terminal variant of betaII spectrin: two
```

```
RT   isoforms of betaII spectrin have distinct intracellular locations and
RT   activities.";
RL   J. Cell Sci. 113:2023-2034(2000).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 173-280.
RX   MEDLINE=97307247; PubMed=9164454;
RA   Carugo K.D., Banuelos S., Saraste M.;
RT   "Crystal structure of a calponin homology domain.";
RL   Nat. Struct. Biol. 4:175-179(1997).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 173-281.
RX   MEDLINE=99036861; PubMed=9817844;
RA   Banuelos S., Saraste M., Carugo K.D.;
RT   "Structural comparisons of calponin homology domains: implications
RT   for actin binding.";
RL   Structure 6:1419-1431(1998).
CC   -!- FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION,
CC       INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS
CC       THUS CANDIDATE FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE
CC       CYTOSKELETON AT THE MEMBRANE.
CC   -!- SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS ARE
CC       CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO TETRAMERS.
CC       The short form can not bind to the axonal protein fodaxin.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
CC       short form; are produced by alternative splicing.
CC   -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC   -!- SIMILARITY: CONTAINS 17 SPECTRIN REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M96803; AAA60580.1; -.
DR   EMBL; S65762; AAB28324.1; -.
DR   EMBL; AJ005694; CAA06678.1; -.
DR   EMBL; AJ238723; CAB91088.1; -.
DR   PDB; 1AA2; 04-FEB-98.
DR   PDB; 1BKR; 26-FEB-99.
DR   MIM; 182790; -.
DR   InterPro; IPR001589; Actinin_act_bind.
DR   InterPro; IPR001715; Calponin_hom.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR001605; Spectrin_PH.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00435; spectrin; 17.
DR   PRINTS; PR00683; SPECTRINPH.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00150; SPEC; 16.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00020; ACTININ_2; 1.
DR   PROSITE; PS50021; CH; 2.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
KW   Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein;
KW   Alternative splicing; 3D-structure.
FT   DOMAIN        1    275       ACTIN-BINDING (BY SIMILARITY).
FT   DOMAIN       54    158       CH 1.
FT   DOMAIN      173    275       CH 2.
FT   REPEAT      276    384       SPECTRIN 1.
FT   REPEAT      385    498       SPECTRIN 2.
FT   REPEAT      499    608       SPECTRIN 3.
FT   REPEAT      609    714       SPECTRIN 4.
FT   REPEAT      715    819       SPECTRIN 5.
FT   REPEAT      820    925       SPECTRIN 6.
FT   REPEAT      926   1032       SPECTRIN 7.
FT   REPEAT     1033   1139       SPECTRIN 8.
FT   REPEAT     1140   1245       SPECTRIN 9.
FT   REPEAT     1246   1350       SPECTRIN 10.
FT   REPEAT     1351   1462       SPECTRIN 11.
FT   REPEAT     1463   1562       SPECTRIN 12.
FT   REPEAT     1563   1668       SPECTRIN 13.
FT   REPEAT     1669   1775       SPECTRIN 14.
FT   REPEAT     1776   1881       SPECTRIN 15.
FT   REPEAT     1882   1987       SPECTRIN 16.
FT   REPEAT     1988   2133       SPECTRIN 17.
FT   DOMAIN     2197   2307       PH.
FT   VARSPLIC   2141   2168       MAETVDTSEMVNGATEQRTSSKESSPIP -> VSYRSQTYQ
FT                                NYKNFNSRRTASDQPWSGL (IN SHORT ISOFORM).
FT   VARSPLIC   2169   2364       MISSING (IN SHORT ISOFORM).
FT   CONFLICT   1411   1411       H -> D (IN REF. 2).
SQ   SEQUENCE   2364 AA;  274629 MW;  0122DB3DF25872CC CRC64;

  Query Match             3.0%;  Score 130.5;  DB 1;  Length 2364;
  Best Local Similarity   17.4%;  Pred. No. 7.7;
  Matches  174;  Conservative  156;  Mismatches  286;  Indels  383;  Gaps   47;

Qy    118 IQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEI 177
          |:   :|  :|:| |     :::               | ||::  :  |
Db    298 IETEKMIEKYESLASDLLEWIEQ-------------TIIILNNRKFANS---------- 333

Qy    178 ERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQS-LASNGWKLES 236
            |||:|  |           | :   |      ||  :| ||  ||| : :|  |:
Db    334 ---LVGVQQQL--------QAFNTYRTVEKPPKFTEKGNLEVLLFTIQSKMRANNQKV-Y 381

Qy    237 VPRPHLSFEAQLVA-----------------GKFHELRPIKCMEQPSPPSDHSRAY--- 275
          :||    | :|::                  :||   : :|| :   |   |
Db    382 MPR-----EGKLISDINKAWERLEKAEHERELALRNELIRQEKLEQLARRFDRKAAMRET 436

Qy    276 ----------------------SGKQKHDA----LTRYPQRIRRLNIFPANKME-----D 304
                               :  :||:|    :  | :|:: : :  | ::|     |
Db    437 WLSENQRLVSQDNFGFDLPAVEAATKKHEAIETDIAAYEERVQAV-VAVARELEAENYHD 495

Qy    305 VQPI----DRFV-VEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLF------ 353
          :: |    |  : : ||||::|      |::       | :    : :  | |:
Db    496 IKRITARKDNVIRLWEYLLELL----RARRQ------RLEMNLGLQKIFQEMLYIMDWMD 545

Qy    354 -SQILLLPQPPFKTLY--------YTLVIMDL--------CKALPGAF----------- 384
           ::|:| |   | |          :|||  |:           |    |
Db    546 EMKVLVLSQDYGKHLLGVEDLLQKHTLVEADIGIQAERVRGVNASAQKFATDGEGYKPCD 605

Qy    385 PAVVAGAVRAL---FEKISDLDMESRTRL----ILWFSHHLSNFQFIWPWEE---WAFVL 434
          | |:   |  :   :::: |  | | ||     ||        :| |   |   |
Db    606 PQVIRDRVAHMEFCYQELCQLAAERRARLEESRRLW-------KFFWEMAEEEGW---- 653

Qy    435 DLPKWAPKRVFVQEILQREVRLSYWD------KIKQSIENATALEELLPPKAGPNFMYSL 488
                    | ::|  ||  |       : : :    | |: :  ::| :|  ::
Db    654 -------------IREKEKILSSDDYGKDLTSVMRLLSKHRAFEDEMSGRSG-HFEQAI 698

Qy    489 EEGK----------EKTEE-----------QQLSAELSRKVKE-------KQTARDMIV 519
          :||:          ||  |            :||||   ::::|       :  | |:
Db    699 KEGEDMIAEEHFGSEKIRERIIYIREQWANLEQLSAIRKKRLEEASLLHQFQADADDIDA 758

Qy    520 WIEETIYPVHGFEV-----TLTIVVQTLLDIGSKSFTHLVT------------------ 555
          |: : :  |   :|     :   :|:  |: :   :  |
Db    759 WMLDILKIVSSSDVGHDEYSTQSLVKKHKDVAEEIANYRPTLDTLHEQASALPQEHAESP 818

Qy    556 --------VLERYGQV---------------FSKLCPDNDKQVMLLSQVSTYWKNNVQM 591
                 : ||| :|                 |:  : |   : : :    | ||:|:
Db    819 DVRGRLSGIEERYKEVAELTRLRKQALQDTLALYKMFSEADACELWIDE-KEQWLNNMQI 877

Qy    592 -------------------------TAVAIDRMMGYRLV-------------SNQAIVRW 613
                                  : ||:   :  :|:                ::   ||
Db    878 PEKLEDLEVIQHRFESLEPEMNNQASRVAVVNQIARQLMHSGHPSEKEIKAQQDKLNTRW 937

Qy    614 ------------VFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRK------DISNI 654
                      : |  ::  :|:           |:| : |  :  |      |: |
Db    938 SQFRELVDRKKDALLSALSIQNYHLE----------CNETKSWIREKTKVIESTQDLGND 987
```

```
Qy     655 TKNVLVAEKASANARVELEAAESKLSLV--EGEPVLGENPAKMKRLKSTVEKTGEAELSL 712
           |:  ::       :| | |:||| :  | | :  |:| : : : | : :  :    :
Db     988 LAGVMALQRKLTGMERDLVAIEAKLSDLQKEAEKLESEHPDQAQAILSRLAEISDVWEEM 1047

Qy     713 RESLEAKEALLNRALSETEVLLLLL--FQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDK 770
           : :|: :|| |  | |: :  |  |  |||:|   :  :                 | :|
Db    1048 KTTLKNREASLGEA-SKLQQFLRDLDDFQSWLSRTQTAI---------------ASEDM 1090

Qy     771 PSAM--------------DVDSENGNPKKSCEVGEREQWCLSTLGYLTA---FTRQYAS 812
           |: :              ::|:   : :|  ::||     : | |   |   | ||
Db    1091 PNTLTEAEKLLTQHENIKNEIDNYEEDYQKMRDMGE-----MVTQGQTDAQYMFLRQRLQ 1145

Qy     813 EI---WPHMEKL---ESEVFSGEDVHPLFLQAISSALQF 845
           :   |  : |:      : |    :  ||:    |  |
Db    1146 ALDTGWNELHKMWENRQNLLSQSHAYQQFLRDTKQAEAF 1184
```

RESULT 12
ACF7_HUMAN
ID ACF7_HUMAN STANDARD; PRT; 5430 AA.
AC Q9UPN3; Q9UKP0; Q9ULG9; Q9H540; O75053;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Actin cross-linking family protein 7 (Macrophin) (Trabeculin-alpha)
DE (620 kDa actin-binding protein) (ABP620).
GN ACF7 OR ABP620 OR KIAA0465 OR KIAA1251.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20001959; PubMed=10529403;
RA Okuda T., Matsuda S., Nakatsugawa S., Ichigotani Y., Iwahashi N.,
RA Takahashi M., Ishigaki T., Hamaguchi M.;
RT "Molecular cloning of macrophin, a human homologue of Drosophila
RT kakapo with a close structural similarity to plectin and dystrophin.";
RL Biochem. Biophys. Res. Commun. 264:568-574(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20026884; PubMed=10559237;
RA Sun Y., Zhang J., Kraeft S.-K., Auclair D., Chang M.-S., Liu Y.,
RA Sutherland R., Salgia R., Griffin J.D., Ferland L.H., Chen L.B.;
RT "Molecular cloning and characterization of human trabeculin-alpha, a
RT giant protein defining a new family of actin-binding proteins.";
RL J. Biol. Chem. 274:33522-33530(1999).
RN [3]
RP SEQUENCE OF 868-2350 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirosawa M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
RN [4]
RP SEQUENCE OF 1544-5057 FROM N.A.
RA Corby N.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE OF 3734-5430 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98116662; PubMed=9455484;
RA Seki N., Ohira M., Nagase T., Ishikawa K.-I., Miyajima N.,
RA Nakajima D., Nomura N., Ohara O.;
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain.";
RL DNA Res. 4:345-349(1997).
CC -!- FUNCTION: F-ACTIN-BINDING PROTEIN WHICH MAY PLAY A ROLE IN CROSS-
CC   LINKING ACTIN TO OTHER CYTOSKELETAL PROTEINS. ALSO BINDS TO
CC   MICROTUBULES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED.
CC -!- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 37 SPECTRIN REPEATS.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AB029290; BAA83821.1; -.
DR EMBL; AF141968; AAF06360.1; -.
DR EMBL; AB033077; BAA86565.1; -.
DR EMBL; AL137853; CAC15920.1; -.
DR EMBL; AB007934; BAA32310.1; -.
DR HSSP; Q01082; 1BKR.
DR InterPro; IPR001589; Actinin_act_bind.
DR InterPro; IPR001715; Calponin_hom.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR003108; GAS2.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00036; efhand; 2.
DR Pfam; PF02187; GAS2; 1.
DR Pfam; PF00435; spectrin; 35.
DR SMART; SM00033; CH; 2.
DR SMART; SM00054; EFh; 2.
DR SMART; SM00243; GAS2; 1.
DR SMART; SM00150; SPEC; 35.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR PROSITE; PS50021; CH; 2.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS50002; SH3; FALSE_NEG.
KW Actin-binding; Cytoskeleton; Calcium-binding; Repeat; SH3 domain.

FT	DOMAIN	1	295	ACTIN-BINDING.
FT	DOMAIN	78	181	CH 1.
FT	DOMAIN	194	295	CH 2.
FT	REPEAT	314	355	SPECTRIN 1.
FT	REPEAT	591	623	SPECTRIN 2.
FT	REPEAT	680	784	SPECTRIN 3.
FT	REPEAT	786	800	SPECTRIN 4.
FT	DOMAIN	871	923	SH3.
FT	REPEAT	1250	1272	SPECTRIN 5.
FT	REPEAT	1287	1342	SPECTRIN 6.
FT	REPEAT	1455	1534	SPECTRIN 7.
FT	REPEAT	1547	1659	SPECTRIN 8.
FT	REPEAT	1815	1891	SPECTRIN 9.
FT	REPEAT	1932	2042	SPECTRIN 10.
FT	REPEAT	2260	2280	SPECTRIN 11.
FT	REPEAT	2372	2395	SPECTRIN 12.
FT	REPEAT	2398	2507	SPECTRIN 13.
FT	REPEAT	2510	2618	SPECTRIN 14.
FT	REPEAT	2621	2728	SPECTRIN 15.
FT	REPEAT	2731	2838	SPECTRIN 16.
FT	REPEAT	2841	2945	SPECTRIN 17.
FT	REPEAT	2987	3024	SPECTRIN 18.
FT	REPEAT	3136	3163	SPECTRIN 19.
FT	REPEAT	3187	3274	SPECTRIN 20.
FT	REPEAT	3277	3383	SPECTRIN 21.
FT	REPEAT	3386	3492	SPECTRIN 22.
FT	REPEAT	3495	3601	SPECTRIN 23.
FT	REPEAT	3604	3673	SPECTRIN 24.
FT	REPEAT	3713	3819	SPECTRIN 25.
FT	REPEAT	3832	3927	SPECTRIN 26.

```
FT   REPEAT      3982    4043       SPECTRIN 27.
FT   REPEAT      4046    4152       SPECTRIN 28.
FT   REPEAT      4155    4262       SPECTRIN 29.
FT   REPEAT      4265    4371       SPECTRIN 30.
FT   REPEAT      4374    4481       SPECTRIN 31.
FT   REPEAT      4484    4590       SPECTRIN 32.
FT   REPEAT      4593    4700       SPECTRIN 33.
FT   REPEAT      4707    4808       SPECTRIN 34.
FT   REPEAT      4811    4917       SPECTRIN 35.
FT   REPEAT      4920    4985       SPECTRIN 36.
FT   REPEAT      5030    5054       SPECTRIN 37.
FT   CA_BIND     5096    5107       EF-HAND 1 (POTENTIAL).
FT   CA_BIND     5132    5143       EF-HAND 2 (POTENTIAL).
FT   DOMAIN      5276    5283       POLY-SER.
FT   DOMAIN      5355    5370       4 X 4 AA TANDEM REPEATS OF [GS]-S-R-[AR].
FT   CONFLICT       1      72       MSSSDEETLSERSCRSERSCRSERSYRSERSGSLSPCPPGD
FT                                  TLPWNLPLHEQKKRKSQDSVLDPAERAVVRV -> MFPVLW
FT                                  AGIPGRDVGSLQPLPPGFKQFCTSASRVAVI (IN REF.
FT                                  2).
FT   CONFLICT     575     594       VAISSSEDEGNLRFVYELLS -> GPSAPLKMKAISDLCMN
FT                                  YCL (IN REF. 1).
FT   CONFLICT    1487    1487       A -> T (IN REF. 2 AND 3).
FT   CONFLICT    1963    1963       V -> A (IN REF. 1).
FT   CONFLICT    2052    2052       E -> D (IN REF. 1).
FT   CONFLICT    2083    2083       E -> K (IN REF. 2).
FT   CONFLICT    2290    2290       M -> V (IN REF. 2).
FT   CONFLICT    2321    2321       C -> Y (IN REF. 1).
FT   CONFLICT    2343    2363       MISSING (IN REF. 2).
FT   CONFLICT    2344    2350       SILPSVG -> EYRLFKI (IN REF. 3).
FT   CONFLICT    2523    2523       Q -> R (IN REF. 2 AND 4).
FT   CONFLICT    4670    4670       S -> T (IN REF. 2).
FT   CONFLICT    4833    4833       MISSING (IN REF. 2).
SQ   SEQUENCE   5430 AA;  620346 MW;  91ADB7F7580B440B CRC64;

  Query Match             3.0%;  Score 130;  DB 1;  Length 5430;
  Best Local Similarity   19.3%;  Pred. No. 27;
  Matches  162;  Conservative  134;  Mismatches  339;  Indels  204;  Gaps   36;

Qy      52 QCAEQL-PHKIP---LYGTLIGLLNLENED-----FVQKLVESVHANFQVALDSGNCNSI 102
           |  |:|  |:::|   |       :: :| |       ||:  :|: ::||         |:
Db    1979 QLQEELAEHQVPVEKLQKVARDIMEIEGEPAPDHRHVQETTDSILSHFQ---------SL 2029

Qy     103 RILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSL 162
             |   :|||   : |  |:    |:|| |       :|                :|||
Db    2030 SYSLAERSSLLQKAIAQSQSVQESLESLLQSIGEVEQNLEGKQ------------VSSL 2076

Qy     163 PWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLD 222
             |  : |     :::: :   :: |   :   |    |     :  ::   :  : ::
Db    2077 SSGVIQEALATNMKLKQDIARQKSSLEATREMVT---RFMETADSTTAAVLQGKLAEVSQ 2133

Qy     223 RIQSLASNGWKLES-----VPRPHLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSG 277
           | : |     : ||     :|:  : ||   ::|| :      || |          ||
Db    2134 RFEQLCLQQQEKESSLKKLLPQAEM-FEH--LSGKLQQF-----MENKS-----RMLASG 2180

Qy     278 KQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVE----EYLLDVLFY---LNGCRKE 330
             |    :| : |:|: ||:   :: |::  ::  | |     : ||:  :   :   ||:
Db    2181 NQPDQDITHFFQQIQELNLEMEDQQENLDTLEHLVTELSSCGFALDLCQHQDRVQNLRKD 2240

Qy     331 CASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAG 390
                 |  | :     | :   |:       |: |  |       |   |: |
Db    2241 ----FTELQKTVKEREKDASSCQEQL-----DEFRKLVRT--FQKWLKETEGSIPPTET- 2288

Qy     391 AVRALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEIL 450
                  :|  ::| :          || :             |  || |   |:||
Db    2289 -------SMSAKELEKQI--------EHLKSL-------------LDDWASKGTLVEEIN 2320

Qy     451 QREVRL-----------------SYWDKIKQSIENATA------LEELLPPKAGPNFMYS 487
            :  |                  |    :  |: :         | |:    :  |  |
Db    2321 CKGTSLENLIMEITAPDSQGKTGSILPSVGSSVGSVNGYHTCKDLTEIQCDMSDVNLKYE 2380

Qy     488 LEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGS 547
                |     :: | | |:|  :  : |  :: |:|          :   |:::: : |

Db    2381 KLGGVLHERQESLQAILNRMEEVHKEANSVLQWLE----------SKEEVLKSMDAMSS 2429

Qy     548 KSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSN 607
             : |  |       : |      |  ::  : :            |:|
Db    2430 PTKTETVKAQAESNKAFLAELEQNSPKIQKVKE-----------ALA---------- 2465

Qy     608 QAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASAN 667
             :| :  | |         : |: :   ||  : | :::      : :   : | ||
Db    2466 GLLVTYPNSQE--------AENWKKIQEELNSRWERATEV-----TVARQRQLEESASHL 2512

Qy     668 ARVELEAAESKLS--LVEGEPVLGE------NPAKMKRLKSTVE---KTGEAELSLRESL 716
           |    :|||||:|   |:| | ::|       :|  :   |  |:   |  ||     | |
Db    2513 AC--FQAAESQLQPWLMEKELMMGVLGPLSIDPNMLNAQKQQVQFMLKEFEARRQQHEQL 2570

Qy     717 -EAKEALL----NRALSETEVLLLLLFQSFLGVLKER---LPDPTKVRSVQDLKSIGAED 768
            || : :|    : :|| ::|      |  |  : ::   | |    || |  ::|
Db    2571 NEAAQGILTGPGDVSLSTSQV------QKELQSINQKWVELTDKLNSRSSQIDQAIVKST 2624

Qy     769 DKPSAMDVDSENGNPKKSCEVGER--EQWCLSTLGYLTAFTRQYASEIWPHMEKLESEV 825
             :   ||      |   ||:|   |  :||         :  |||   :|:|: ||
Db    2625 QYQELLQDLSE-----KVRAVGQRLSVQSAISTQPEAVKQQLEETSEIRSDLEQLDHEV 2678


RESULT  13
RA50_YEAST
ID   RA50_YEAST     STANDARD;      PRT;  1312 AA.
AC   P12753;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA repair protein RAD50 (153 kDa protein).
GN   RAD50 OR YNL250W OR N0872.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RE821;
RX   MEDLINE=89276917; PubMed=2659437;
RA   Alani E., Subbiah S., Kleckner N.;
RT   "The yeast RAD50 gene encodes a predicted 153-kD protein containing a
RT   purine nucleotide-binding domain and two large heptad-repeat
RT   regions.";
RL   Genetics 122:47-57(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE=97377992; PubMed=9234673;
RA   Sen-Gupta M., Gueldener U., Beinhauer J.D., Fiedler T.A.,
RA   Hegemann J.H.;
RT   "Sequence analysis of the 33 kb long region between ORC5 and SUI1
RT   from the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RL   Yeast 13:849-860(1997).
CC   -!- FUNCTION: INVOLVED IN DNA DOUBLE-STRAND BREAK REPAIR (DSBR). THE
CC       RAD50/MRE11 COMPLEX POSSESSES SINGLE-STRAND ENDONUCLEASE ACTIVITY
CC       AND ATP-DEPENDENT DOUBLE-STRAND-SPECIFIC EXONUCLEASE ACTIVITY.
CC       RAD50 PROVIDES AN ATP-DEPENDENT CONTROL OF MRE11 BY UNWINDING
CC       AND/OR REPOSITIONING DNA ENDS INTO THE MRE11 ACTIVE SITE.
CC   -!- SUBUNIT: FORMS A COMPLEX WITH MRE11.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X14814; CAA32919.1; -.
DR   EMBL; X96722; CAA65494.1; -.
DR   EMBL; Z71526; CAA96157.1; -.
DR   PIR; S05808; BWBYDL.
```

```
DR   SGD; S0005194; RAD50.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
KW   DNA repair; Hydrolase; ATP-binding; Coiled coil; Meiosis.
FT   NP_BIND      34     41       ATP (POTENTIAL).
FT   DOMAIN      185    347       COILED COIL (POTENTIAL).
FT   DOMAIN      403    558       COILED COIL (POTENTIAL).
FT   DOMAIN      617    672       COILED COIL (POTENTIAL).
FT   DOMAIN      734   1108       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1312 AA;  152568 MW;  58A0AA173AC5677E CRC64;


  Query Match             3.0%;  Score 129.5;  DB 1;  Length 1312;
  Best Local Similarity   18.0%;  Pred. No. 3.8;
  Matches  164;  Conservative  157;  Mismatches  309;  Indels  283;  Gaps   38;

Qy      68 IGLLNLENEDFVQKLVESVHANF-QVALDSGNCNSIRILLRFMTSLLCSKVIQPASLIVV 126
           | :|:|   |     :::: ||| :| :|  |        || : : : |   : :||  :
Db     292 IDILDLSKPD-----LQNLLANFSKVLMDKNN------QLRDLETDISSLKDRQSSLQSL 340

Qy     127 FETL------LSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERV 180
             :|      | :   | :: : : |   :|  :    ||::    |::| ||  |: :
Db     341 SNSLIRRQGELEAGKETYEKNRNHLSSLKEAFQHKFQGLSNIE--NSDMA-QVNHEMSQ- 396

Qy     181 LVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLE--SVP 238
           :|::|      |  ::     :|   :  |:   :| | |:|:  :   ||
Db     397 ---FKAFIS---QDLTDTID-----QFAKDIQLKE--TNLSDLIKSITVDSQNLEYNKKD 443

Qy     239 RPHLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGK----QKHDALTRYPQRIRRL 294
           |  |  ::: :| |    : :   :  :   :: : |  |    :  : : : :  |:|
Db     444 RSKLIHDSEELAEKLKSFKSLSTQDSLNHELENLKTYKEKLQSWESENIIPKLNQKIEEK 503

Qy     295 N---IFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANL------------- 338
           |   |   |::|  |  || :      |:   |   :|   : :  |
Db     504 NNEMIILENQIEKFQ--DRIMKTNQQADLYAKLGLIKKSINTKLDELQKITEKLQNDSRI 561

Qy     339 ----PVTFRYEYLMAETLFSQILLLPQPPF----KTLY-----YTLVIMDLCKALPGAFP 385
           |:|  ::    |  | :: :  |       | ::     ||  : :|
Db     562 RQVFPLTQEFQRADLEMDFQKLFINMQKNIAINNKKMHELDRRYTNALYNL--------- 612

Qy     386 AVVAGAVRALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVF 445
                 ||    : :|: ::|   | :|     |   :|:  ||:
Db     613 --------NTIEKDLQDNQKSKEKVIQLLSENLPEDCTI---DEYNDVLE---------- 651

Qy     446 VQEILQREVRLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKE----------KT 495
               |  |||          |||| |   :    |    |  |          |
Db     652 -------ETELSY----------KTALENLKMHQTTLEFNRKALEIAERDSCCYLCSRKF 694

Qy     496 EEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVT 555
           | :   ::| :::| |  |                  || ||   ||       | : | :
Db     695 ENESFKSKLLQELKTKTDA--------------NFEKTLKDTVQ-----NEKEYLHSLR 734

Qy     556 VLERY----GQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIV 611
           :||::      :  |:  || :: :  ::  |    : ::  : :|     :|   : :
Db     735 LLEKHIITLNSINEKI--DNSQKCLEKAKEETKTSKS-KLDELEVD---STKLKDEKELA 788

Qy     612 RWVFSP----------------------ENVDQFHVSD---QPWEILGNALNKTYNRIS 645
                |                      | :  :: |:   |  : | :   |  : :
Db     789 ESEIRPLIEKFTYLEKELKDLENSSKTISEELSIYNTSEDGIQTVDELRDQQRKMNDSLR 848

Qy     646 DLRKDISNITKNVLVAEKASANARV-------ELEAAESKLSLVEGEPVLGENPAKMKRL 698
           :||| ||::   :   ||  |:|:       ||  :| : || : : :      :| : :
Db     849 ELRKTISDL--QMEKDEKVRENSRMINLIKEKELTVSEIESSLTQKQNIDDSIRSKRENI 906

Qy     699 KSTVEKTGEAE---LSLRESLEAKEALLNRALSETEV------------LLLLLFQSFL 742
                : | |   :||:   :  :::|::  :| ::              |:  ||:
Db     907 NDIDSRVKELEARIISLKNKKDEAQSVLDKVKNERDIQVRNKQKTVADINRLIDRFQTIY 966

Qy     743 G--------------------------VLKERLPDPTKVRSVQDLKSIGAEDDKPSAM 774
                                       |||:|          ||||    ::|
Db     967 NEVVDFEAKGFDELQTTIKELELNKAQMLELKEQL----------DLKSNEVNEEKRKLA 1016

Qy     775 DVDSENGNPKKSCEV----------------------GEREQWCLSTLGYLTAFTRQYA 811
```

```
           | ::|  | |:: |:                        ||:::   :|   | |
Db    1017 DSNNEEKNLKQNLELIELKSQLQHIESEISRLDVQNAEAERDKYQEESLRLRTRF----- 1071

Qy     812 SEIWPHMEKLESE 824
                 ||| ||
Db    1072 -------EKLSSE 1077


RESULT  14
YD86_SCHPO
ID   YD86_SCHPO     STANDARD;      PRT;  1957 AA.
AC   Q10411;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 222.8 kDa protein C1F3.06C in chromosome I.
GN   SPAC1F3.06C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z70690; CAA94624.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1957 AA;  222785 MW;  3F480CA06171D9DA CRC64;


  Query Match             3.0%;  Score 129.5;  DB 1;  Length 1957;
  Best Local Similarity   20.7%;  Pred. No. 6.8;
  Matches  191;  Conservative  125;  Mismatches  287;  Indels  321;  Gaps   45;

Qy      38 EIERSGDQVLPFLLQCAEQLPHKIPLYGTLIGLL-------NLENEDFVQKL--VESVHA 88
           |:||:       ||   |:|   :   |  |  |          ||:|   ||  | | :
Db    1003 EVERNA------LLASNERLMDDLKNNGENIASLQTEIEKKRAENDDLQSKLSVVSSEYE 1056

Qy      89 NFQVALDSGN------CNSIRILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTVDEEK 142
           |  :     |       | :: : : :  ||  |  :   |  . | | |      :|
Db    1057 NLLLISSQTNKSLEDKTNQLKYIEKNVQKLLDEKDQRNVEL----EELTSKYGKLGEEN- 1111

Qy     143 GNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFF 202
                                        |: ||:          |::|| |
Db    1112 ---------------------------AQIKDEL----------LALRKKSKKQ----- 1128

Qy     203 HN--GEFESSLAEK-DFVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELRPI 259
           |:    |   | || | :| | :    |      : |: : : : || ||  :      :
Db    1129 HDLCANFVDDLKEKSDALEQLTNEKNEL------IVSLEQSNSNNEA-LVEERSDLANRL 1181

Qy     260 KCMEQPSPPSD-------------HSRAYSGKQKHDAL-TRYPQRIRRLNIFPANKMEDV 305
             |::   ||             :   : |:  |:| |:|           |
Db    1182 SDMKKSLSDSDNVISVIRSDLVRVNDELDTLKKDKDSLSTQY--------------SEVC 1227

Qy     306 QPIDRFVVEEYLLDVLFYLNGC----------RKECASYMANLPVTFRYEYLMAETLFS 354
           |  |         |:|  | ||           |: |     ::|||:   | |    :|:
Db    1228 QDRD---------DLLDSLKGCEESFNKYAVSLRELCTKSEIDVPVS---EILDDNFVFN 1275

Qy     355 QILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRAL------FEKISDLDMESRT 408
                                   | |  :   | :|      | :::   ||
Db    1276 -------------------------AGNFSELSRLTVLSLENYLDAFNQVNFKKMELDN 1309
```

```
Qy     409 RLILWFSHHLSNFQFIWPWEEWAFVLDLPK-------WAPKRVFVQEILQREVRLSYWDK 461
           ||        :: :|         | || |        |  :|  ::: | ::   ::  |
Db    1310 RLT------TTDAEFT------KVVADLEKLQHEHDDWLIQRGDLEKAL-KDSEKNFLRK 1356

Qy     462 IKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTA------- 514
             :  ||              ::|||||||:|:::   |||| :::: | |
Db    1357 EAEMTEN---------------IHSLEEGKEETKKE--IAELSSRLEDNQLATNKLKNQ 1398

Qy     515 -----------------RDMIVWIEETIYPVHGFEVTLTIV---VQTLLDIGSKSFTHL 553
                            :|: :||::      | :|      :: :||  |:  : |
Db    1399 LDHLNQEIRLKEDVLKEKESLIISLEESLSNQRQKESSLLDAKNELEHMLDDTSRKNSSL 1458

Qy     554 VTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSN-QAIVR 612
             :|:   : | |   :||   | | |        :: |:         |: | :: ::
Db    1459 ---MEKIESINSSL---DDKSFELASAVE-------KLGALQKLHSESLSLMENIKSQLQ 1505

Qy     613 WVFSPENVDQFHVSDQPWEILGN------ALNKTYNRISDLRKDISNITKNVLVAEKASA 666
                ||:  : :   ||  :       ||   : | || ::|  :  | |:||: ||
Db    1506 EAKEKIQVDESTIQELDHEITASKNNYEGKLNDKDSIIRDLSENIEQL--NNLLAEEKSA 1563

Qy     667 NARVELEAAESKL-------------SLVEGEPVLGENPAKMKRLKSTVEKTGEAELSL 712
             |:  |  ||::              | || |  || :  |   | || |   | :|:
Db    1564 VKRLSTE-KESEILQFNSRLADLEYHKSQVESE--LGRSKLK---LASTTE---ELQLAE 1614

Qy     713 RESLEAKEALLN-----------RALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQD 760
             | |     :|:            :|||    |  |  |   : ||       | |:||
Db    1615 NERLSLTTRMLDLQNQVKDLSNIKDSLSEDLRTLRSLEDSVASLQKECKIKSNTVESLQD 1674

Qy     761 -LKSIGA-----EDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLTAFTRQYASEI 814
             | |: |     ||:   ::|           ::  |:  |    | |
Db    1675 VLTSVQARNAELEDEVSRSVD-----------KIRRRDDRCEHLSGKL----------- 1711

Qy     815 WPHMEKLESEVFSGEDVHPLFLQA 838
               :|| |::   |: |  | :|
Db    1712 ----KKLHSQL---EEQHETFFRA 1728


RESULT  15
MYH3_HUMAN
ID   MYH3_HUMAN     STANDARD;      PRT;  1940 AA.
AC   P11055; Q15492;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, fast skeletal muscle, embryonic (Muscle embryonic
DE   myosin heavy chain) (SMHCE).
GN   MYH3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89263803; PubMed=2726495;
RA   Eller M.S., Stedman H.H., Sylvester J.E., Fertels S.H.,
RA   Rubinstein N.A., Kelly A.M., Sarkar S.;
RT   "Nucleotide sequence of full length human embryonic myosin heavy
RT   chain cDNA.";
RL   Nucleic Acids Res. 17:3591-3592(1989).
RN   [2]
RP   SEQUENCE OF 774-1940 FROM N.A.
RX   MEDLINE=90033298; PubMed=2806546;
RA   Eller M.S., Stedman H.H., Sylvester J.E., Fertels S.H., Wu Q.-L.,
RA   Raychowdhury M.K., Rubinstein N.A., Kelly A.M., Sarkar S.;
RT   "Human embryonic myosin heavy chain cDNA. Interspecies sequence
RT   conservation of the myosin rod, chromosomal locus and isoform
RT   specific transcription of the gene.";
RL   FEBS Lett. 256:21-28(1989).
RN   [3]
RP   SEQUENCE OF 856-1940 FROM N.A.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=90235862; PubMed=1691980;
RA   Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
RA   Arnold H.H.;
RT   "Identification of three developmentally controlled isoforms of human
RT   myosin heavy chains.";
RL   Eur. J. Biochem. 189:55-65(1990).
RN   [4]
RP   SEQUENCE OF 856-1940 FROM N.A.
RX   MEDLINE=89366648; PubMed=2771643;
RA   Karsch-Mizrachi I., Travis M., Blau H., Leinwand L.A.;
RT   "Expression and DNA sequence analysis of a human embryonic skeletal
RT   muscle myosin heavy chain gene.";
RL   Nucleic Acids Res. 17:6167-6179(1989).
CC   -!- FUNCTION: MUSCLE CONTRACTION.
CC   -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC       HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC       AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- DEVELOPMENTAL STAGE: ABUNDANTLY PRESENT IN FETAL SKELETAL MUSCLE
CC       AND NOT PRESENT OR BARELY DETECTABLE IN HEART AND ADULT SKELETAL
CC       MUSCLE.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC       ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC   -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC       SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement.(See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13988; CAA32167.1; -.
DR   EMBL; X13100; CAA31492.1; -.
DR   EMBL; X51593; CAA35942.1; -.
DR   EMBL; X15696; CAA33731.1; -.
DR   PIR; S04090; S04090.
DR   HSSP; P08799; 1MMD.
DR   MIM; 160720; -.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW   Multigene family.
FT   DOMAIN        1    781       MYOSIN HEAD-LIKE.
FT   DOMAIN      782    811       IQ.
FT   DOMAIN      840   1933       COILED COIL (POTENTIAL).
FT   NP_BIND     179    186       ATP (POTENTIAL).
FT   DOMAIN      656    678       ACTIN-BINDING.
FT   DOMAIN      758    772       ACTIN-BINDING.
FT   MOD_RES     130    130       METHYLATION (TRI-) (POTENTIAL).
FT   MOD_RES     696    696       ALKYLATION (SH-1).
FT   MOD_RES     706    706       ALKYLATION (SH-2).
FT   CONFLICT   1331   1331       A -> G (IN REF. 3).
FT   CONFLICT   1391   1392       KK -> QE (IN REF. 1 AND 2).
```

```
FT   CONFLICT    1608    1609       SR -> RA (IN REF. 3).
FT   CONFLICT    1663    1664       RG -> QT (IN REF. 2).
SQ   SEQUENCE   1940 AA;  224035 MW;  43CA58C6A4BA1253 CRC64;


  Query Match             2.9%;  Score 128.5;  DB 1;  Length 1940;
  Best Local Similarity   17.9%;  Pred. No. 7.7;
  Matches  163;  Conservative  146;  Mismatches  307;  Indels  295;  Gaps   41;

Qy      15 GPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHKIPLYGTLIGLLN-- 72
           | || |     | :       : : :       |  : :  :||  |:|     ||:|:
Db     408 GNEYVTKGQTVDQVHHAVNALSKSVYEK--LFLWMVTRINQQLDTKLPRQ-HFIGVLDIA 464

Qy      73 ------------------------------LENEDFVQKLVESVHANFQVALDSGNCN 100
                                           || |:: :: :|    :|   :|   |
Db     465 GFEIFEYNSLEQLCINFTNEKLQQFFNHHMFVLEQEEYKKEGIEWTFIDF--GMDLAAC- 521

Qy     101 SIRILLRFM--TSLLCSKVIQPASLIVVFETLLSSAATTVDEE--KGNPSWQP------- 149
           . | :: : |   |:|  : : | :   |:  |          |:   | |   :|
Db     522 -IELIEKPMGIFSILEEECMFPKATDTSFKNKL------YDQHLGKSNNFQKPKVVKGRA 574

Qy     150 QADFYVICILSSLPWGGSELAEQVPDEIERVLVGI----------QAYLSIRKNSSTSGL 199
           :| | :|    :: :  |   |:  | :   :||:            | :     : ||
Db     575 EAHFSLIHYAGTVDYSVSGWLEKNKDPLNETVVGLYQKSSNRLLAHLYATFATADADSG- 633

Qy     200 NFFHNGEFESSLAEKD----------FVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLV 249
                 :  :|:|           | |:|   : :| :          ||
Db     634 --------KKKVAKKKGSSFQTVSALFRENLNKLMSNLRTT--------HPHF------- 670

Qy     250 AGKFHELRPIKCM---EQPSPPS-DHSRAYSGKQKHDAL-------TRYPQRI------R 292
                 ::|:   |  :| : :||      : :  |        :| ||      :
Db     671 ---------VRCIIPNETKTPGAMEHSLVLHQLRCNGVLEGIRICRKGFPNRILYGDFKQ 721

Qy     293 RLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTF-RYEYLMAET 351
           |  :  |: : : | ||                 :| |   :|:: :   :|::   :
Db     722 RYRVLNASAILEGQFID----------------SKKACEKLLASIDIDHTQYKFGHTKV 764

Qy     352 LFSQILLLPQPPPFKTLYYTLVIM--DLCKALPGAFPAVVAG-AVRALFEKISDLDMESRT 408
            |   ||           ||  |  |    |     ||  |  :|  |:|     |  |
Db     765 FFKAGLL----------GTLEEMRDDRLAKLITRTQAVCRGFLMRVEFQK-----MVQRR 809

Qy     409 RLILWFSHHLSNFQFI--WPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSI 466
             |    ::: :|  :  ||| :  |                          |||  :
Db     810 ESIFCIQYNIRSFMNVKHWPWMKLFF--------------------------KIKPLL 841

Qy     467 ENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIY 526
           ::|   :|          | :::|  :||:::   :|  ||  |::
Db     842 KSAETEKE----------MATMKEEFQKTKDELAKSEAKRKELEEK-------------- 877

Qy     527 PVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWK 586
                                 |||:::           || |: : ::
Db     878 -------------------------LVTLVQE----------KNDLQLQVQAE-----S 896

Qy     587 NNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRISD 646
            |:       |:::  :       |       |: :: :       |     |  :  |:
Db     897 ENLLDAEERCDQLIKAKFQLEAKIKEVTERAEDEEEINAE------LTAKKRKLEDECSE 950

Qy     647 LRKDISNITKNVLVAEKASANARVELEAAESKL-SLVEGEPVLGENPAKMKRLKSTVEKT 705
           |:||| ::   :   ||       |  | |:|: :| |    | |  ||: | |  ::
Db     951 LKKDIDDLELTLAKVEK-------EKHATENKVKNLTEELSGLDETIAKLTREKKALQ-- 1001

Qy     706 GEAELSLRESLEAKEALLNRALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIG 765
            ||      : |:|:|  :| :|::|:  |    :  :  |:  |    |:|   :
Db    1002 -EAHQQALDDLQAEEDKVN-SLNKTKSKL----EQQVEDLESSLEQEKKLRVDLERNKRK 1055

Qy     766 AEDDKPSAMD--VDSENGNPKKSCEVGERE-QWCL--------STLGYLTAFTRQYASEI 814
            | |   | :  :| ||   :    : ::: ::|          |||    | ::   |:
Db    1056 LEGDLKLAQESILDLENDKQQLDERLKKKDFEYCQLQSKVEDEQTLG--LQFQKK-IKEL 1112

Qy     815 WPHMEKLESEV 825
              :|:|| |:
Db    1113 QARIEELEEEI 1123
```

```
Search completed: August 20, 2002, 15:36:29
Job time: 597 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 20, 2002, 14:02:31 ; Search time 51.19 Seconds
(without alignments)
1591.790 Million cell updates/sec

Title: US-09-882-986-2
Perfect score: 4374
Sequence: 1 MSNWKTLLLRIGEKGPEYGT.....EDVHPLFLQAISSALQFPLH 848

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3784	86.5	749	2	C84508	probable cap-bindi
2	813	18.6	790	2	S50082	nuclear cap bindin
3	636.5	14.6	740	2	T15197	hypothetical prote
4	475.5	10.9	780	2	T39057	hypothetical prote
5	279	6.4	858	2	A44919	GCR3 protein - yea
6	166	3.8	879	2	C71083	conserved hypothet
7	163	3.7	1413	2	T26467	hypothetical prote
8	145	3.3	1023	2	AE1643	ATP-dependent dsDN
9	139.5	3.2	3259	1	A56539	giantin - human
10	139	3.2	1388	2	T30335	KLP2 protein - Afr
11	138	3.2	1295	2	T24587	hypothetical prote
12	137.5	3.1	3225	2	I52300	giantin - human
13	137	3.1	1156	2	B70356	chromosome assembl
14	136	3.1	1133	2	T22976	hypothetical prote
15	135	3.1	4151	2	T13734	groovin gene prote
16	134.5	3.1	886	2	H69378	conserved hypothet
17	134.5	3.1	1818	1	S73852	hypothetical prote
18	133.5	3.1	1825	2	T42725	actin binding prot
19	133.5	3.1	1885	2	T30847	actin binding prot
20	133.5	3.1	2033	2	T30849	actin binding prot
21	133.5	3.1	2331	2	T25410	hypothetical prote
22	132	3.0	4063	2	T42993	probable spectrin
23	132	3.0	4101	2	T23630	hypothetical prote
24	131.5	3.0	1313	2	F96673	hypothetical prote
25	131.5	3.0	1864	2	F86378	protein F21J9.12 [
26	131	3.0	1805	1	A64224	hypothetical prote
27	131	3.0	1927	2	G64585	cag pathogenicity
28	131	3.0	2077	2	T43991	large tegument pro
29	131	3.0	2663	1	S28261	centromere protein
30	130.5	3.0	1252	2	A47213	beta-fodrin - huma
31	130.5	3.0	2364	1	A44159	spectrin beta-G ch
32	130	3.0	1298	2	T24480	hypothetical prote
33	130	3.0	2077	2	T44178	large tegument pro
34	129.5	3.0	853	2	T51505	hypothetical prote
35	129.5	3.0	1312	1	BWBYDL	RAD50 protein - ye
36	129.5	3.0	1957	2	T38077	hypothetical coile
37	129	2.9	4131	2	T21085	hypothetical prote
38	128.5	2.9	1048	2	T19045	ras GTPase-activat
39	128.5	2.9	1207	2	T19041	ras GTPase-activat
40	128.5	2.9	1940	1	S04090	myosin heavy chain
41	128	2.9	1819	2	A71928	cag island protein
42	128	2.9	3660	1	S02041	dystrophin, muscle
43	128	2.9	3678	2	S28916	dystrophin - mouse
44	127.5	2.9	284	2	A60607	tropomyosin - fluk
45	127	2.9	964	2	G86491	hypothetical prote

ALIGNMENTS

RESULT 1
C84508
probable cap-binding protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84508
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallc
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Vent
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: C84508
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-749 <STO>
A;Cross-references: GB:AE002093; NID:g4558659; PIDN:AAD22677.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g13540
A;Map position: 2

Query Match 86.5%; Score 3784; DB 2; Length 749;
Best Local Similarity 99.1%; Pred. No. 1.1e-229;
Matches 735; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 MSNWKTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSNWKTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHK 60

Qy      61 IPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 IPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQP 120

Qy     121 ASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERV 180

Qy     181 LVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRP 240

Qy     241 HLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPAN 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 HLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPAN 300

Qy     301 KMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLP 360

Qy     361 QPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHLSN 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    361 QPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHLSN 420

Qy    421 FQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLPPKA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 FQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLPPKA 480

Qy    481 GPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQ 540

Qy    541 TLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMM 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMM 600

Qy    601 GYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLV 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLV 660

Qy    661 AEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKE 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKE 720

Qy    721 ALLNRALSETEVLLLLLFQSFL 742
          ||||||||||||| : | |  :
Db    721 ALLNRALSETEVLYISLMQHMI 742


RESULT   2
S50082
nuclear cap binding protein - human
C;Species: Homo sapiens (man)
C;Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 17-Nov-2000
C;Accession: S50082; A54748
R;Kataoka, N.; Ohno, M.; Kangawa, K.; Tokoro, Y.; Shimura, Y.
Nucleic Acids Res. 22, 3861-3865, 1994
A;Title: Cloning of a complementary DNA encoding an 80 kilodalton nuclear cap binding pr
A;Reference number: S50082; MUID:95023141
A;Accession: S50082
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-790 <KAT>
A;Cross-references: GB:D32002; NID:g1865647; PIDN:BAA06769.1; PID:g577628
R;Izaurralde, E.; Lewis, J.; McGuigan, C.; Jankowska, M.; Darzynkiewicz, E.; Mattaj, I.W
Cell 78, 657-668, 1994
A;Title: A nuclear cap binding protein complex involved in pre-mRNA splicing.
A;Reference number: A54748; MUID:94349369
A;Accession: A54748
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-751,'D',753-790 <IZA>
C;Genetics:
A;Gene: GDB:NCBP
A;Cross-references: GDB:393275; OMIM:600469
A;Map position: 9q22.3-9q22.3
C;Superfamily: human nuclear cap-binding protein
C;Keywords: nucleus; pre-mRNA splicing


  Query Match             18.6%;  Score 813;  DB 2;  Length 790;
  Best Local Similarity   26.6%;  Pred. No. 4.2e-43;
  Matches  231;  Conservative  174;  Mismatches  326;  Indels  136;  Gaps   22;

Qy      5 KTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHKIPLY 64
          ::|: ::|||     ::   : ::|   ||:  ::     ::|  |   |  || |: :|
Db     32 ESLICKVGEK-----SACSLESNLEGLAGVLEADLPNYKSKILRLLCTVARLLPEKLTIY 86

Qy     65 GTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLI 124
           ||:|||| | :|  : ||::    : :| : | |    |:||:: |:   ||   |::
Db     87 TTLVGLLNARNYNFGGEFVEAMIRQLKESLKANNYNEAVYLVRFLSDLVNCHVIAAPSMV 146

Qy    125 VVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGI 184
          :||  :|     | :|:  |  | : |:||  ||||||  | || |:   |::|:
Db    147 AMFENFVS-----VTQEEDVP--QVRRDWYVYAFLSSLPWVGKELYEKKDAEMDRIFANT 199

Qy    185 QAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSF 244
          ::||  |: :    |  :     :    ::::: |  :|| |  : |:   : ||:|:|
Db    200 ESYLKRRQKTHVPMLQVW---TADKPHPQEEYLDCLWAQIQKLKKDRWQERHILRPYLAF 256

Qy    245 EAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMED 304
          :: |     | | |       :||             |::   |: | |:  :  :
Db    257 DSILCEALQHNLPPF------TPP---------PHTEDSVYPMPRVIFRMFDYTDDPEGP 301

Qy    305 VQP----IDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFR--YEYLMAETLFSQILL 358
          | |    ::|||:|| |  ::      || ||: : : |   :    | : | :|:::
Db    302 VMPGSHSVERFVIEENLHCIIKSHWKERKTCAAQLVSYPGKNKIPLNYHIVEVIFAELFQ 361

Qy    359 LPQPPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHL 418
          || ||   : || ::::|||  ||: | |:| |   |: ::  ::     | | ||||||
Db    362 LPAPPHIDVMYTTLLIELCKLQPGSLPQVLAQATEMLYMRLDTMNTTCVDRFINWFSHHL 421

Qy    419 SNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLPP 478
          ||||| | ||:|:  |     :||  ||:|:|:: :|||  :|           :::||
Db    422 SNFQFRWSWEDWSDCLSQDPESPKPKFVREVLEKCMRLSYHQRIL----------DIVPP 471

Qy    479 ------KAGPNFMYSL-EEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYP---- 527
                | |  :|   :|         ::  |:   | | |  ::   :::   |
Db    472 TFSALCPANPTCIYKYGDESSNSLPGHSVALCLAVAFKSKATNDEIFSILKDVPNPNQDD 531

Qy    528 --VHGFE---VTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVS 582
            ||    : : : ||||| : :|||:|  : | :: :||  |   :: :: :|  :
Db    532 DDDEGFSFNPLKIEVFVQTLLHLAAKSFSHSFSALAKFHEVFKTLAESDEGKLHVLRVMF 591

Qy    583 TYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPE---NVDQFHVSDQPWEILGNALNK 639
            |:|: || || :|:|:  ::|   |:  |:|| |   :  :  |    ||||.: : |
Db    592 EVWRNHPQMIAVLVDKMIRTQIVDCAAVANWIFSSELSRDFTRLFV----WEILHSTIRK 647

Qy    640 TYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLK 699
                      : |:||  :|       ||| |: ||:             : ||
Db    648 -------------MNKHVLKIQK-------ELEEAKEKLA------------RQHKRRS 674

Qy    700 STVEKTGEAELS-LRESLEAKEALLNRALSETEVLLLLLFQSFLGVLKERLPDPTKVRSV 758
            ::: : :   | | :|  :  :  | || : | |::|| |: :| | |
Db    675 DDDDRSSDRKDGVLEEQIERLQEKVESAQSEQKNLFLVIFQRFIMILTEHL--------- 725

Qy    759 QDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHM 818
              :  | |  | :                     |  : :  |      |:   |  :|
Db    726 -----VRCETDGTSVL-----------------TPWYKNCIERLQQIFLQHHQIIQQYM 762

Qy    819 EKLESEVFSGE-DVHPLFLQAISSALQ 844
          ||: :|: | | | | :     |||
Db    763 VTLENLLFTAELDPHILAVFQQFCALQ 789


RESULT   3
T15197
hypothetical protein F37E3.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 10-Dec-1999
C;Accession: T15197
R;Wamsley, P.; Bradshaw, H.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid F37E3.
A;Reference number: Z18306
A;Accession: T15197
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-740 <WAM>
A;Cross-references: EMBL:AF003132; NID:g2088676; PID:g2088677; PIDN:AAB54133.1; GSPD
A;Experimental source: strain Bristol N2; clone F37E3
C;Genetics:
A;Gene: CESP:F37E3.1
A;Map position: 1
A;Introns: 26/3; 198/3; 288/3; 363/1; 402/2; 428/3; 480/1; 548/2; 615/2; 662/3; 703/
C;Superfamily: human nuclear cap-binding protein
```

```
  Query Match             14.6%;  Score 636.5;  DB 2;  Length 740;
  Best Local Similarity   24.4%;  Pred. No. 4.7e-32;
  Matches  182;  Conservative  163;  Mismatches  322;  Indels   79;  Gaps   20;

Qy      11 IGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHKIPLYGTLIGL 70
           ||: |   |:| :   :::     :  ::|:    ::  :  ||  ||::: :| ||:|| 
Db      21 IGKVGENTGSSIEC--NLDKLTAFLHDDLEKYRASIIDIIAGCAIYLPNRVTVYTTLVGL 78

Qy      71 LNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLIVVFETL 130
           || :| :|   :|| : :  |  |        : |  |:  |  | |:   |:    |: 
Db      79 LNSKNFNFGGDVVEKLISEQQDLLSKQKYQEAQNLAIFLCDLGNSGVLTAQSIGEYLESF 138

Qy     131 LSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGIQAYLSI 190
           :::|      |:  |  | : |:|:  :|  ||| | || |: |:::| :   |  || : 
Db     139 IAAAF-----EENMP--QVRNDYYIQTVLRCLPWIGKELTEKAPEQMENIGEAIGKYLEL 191

Qy     191 RKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLVA 250
           |  :  : |  :  |  :    ::|::|  |  :|::|  :  |    :|| :  ||  | 
Db     192 RNKNHVALLQVWREGSTDQK--QEDYLESLSAQIEALRNADWVENHIPRHYSGFETTLQD 249

Qy     251 GKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRL------NIFPANKMED 304
           | |   :  |  |                | :  ||  : ||      : | :  : 
Db     250 ALQHNLPSFQSPEHTS--------------DMIYPYPLVVFRLFQDADCSAFSSKPLPG 294

Qy     305 VQPIDRFVVEEYLLDVL----FYLNGCRKECASYM---ANLPVTFRYEYLMAETLFSQIL 357
           ||||: |  :  ::    |    | :|  ::     ::|: |    |: ||:| |:| 
Db     295 DSSIDRFLFEGEIAWIIEKNQFNRKACARELLAFAEENPSVPIGF----LIFETIFGQML 350

Qy     358 LLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHH 417
           ||  |:  :::  ::::| |  |  :| ::   |  :: :   :       |:: ||| | 
Db     351 RLPHAPYPAIFHCSLVLELLKLKPDDYPQILVQTVECIYRRADSMQPVCIDRMVDWFSFH 410

Qy     418 LSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLP 477
           |||||: : | :|   |:   ::  ::||:|:::  |   ::||  ::        :: | 
Db     411 LSNFQYRYTWTDWKDCLNKDAFSGSQIFVREVIEKCRRFGSYEKIIAALPQ--DFVKIHP 468

Qy     478 PKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWI------EETIYPVHGF 531
           |   | ::|  | |   | :   ::  :|:| |  :  :        :|  | :: | 
Db     469 --CSPEVRYLIDE--EDTALVQRAETFTQMFQERQPAEAFLNELKSNDENDELPYNINEF 524

Qy     532 EVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCPDND-KQVMLLSQVSTYWKNNVQ 590
           :  |  :| : ||:::|  : | ||      :|  ::  |  ||  : : || | | 
Db     525 ----GLFVMVMLKMASKTYSHNFSALFRYQTTLKTVCDASELYQEKLLETLYSCWKTNQQ 580

Qy     591 MTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPW--EILGNALNKTYNRISDLR 648
           |  :  |:::  :::   |:| |:|  | : | |  |: |  |:|  || |   :|: : 
Db     581 MLMILTDKLLKMQVIDCSAVVGWLFD-EKMWQEH--DRQWLFEVLNQALEKLTRQINVVE 637

Qy     649 KDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEA 708
           |||  :|:          : |   | :  || : : | |         | :: |  :| 
Db     638 KDIKELTEK-------TENKIKEEDDEESDIKMDEDE-------TKEEKFKQDLEDLENN 683

Qy     709 ELSLRESLEAKEALLNRALSETEVLL 734
           :  |   :  :: | |  |   | || 
Db     684 KEKLERMVTFQKGLFNDFLIHAETLL 709


RESULT   4
T39057
hypothetical protein SPAC6G10.07 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T39057
R;Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z21824
A;Accession: T39057
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-780 <OLI>
A;Cross-references: EMBL:Z98603; NID:e1059110; PIDN:CAB11293.1; GSPDB:GN00066; SPDB:SPAC
A;Experimental source: strain 972h-; cosmid c6G10
C;Genetics:
A;Gene: SPDB:SPAC6G10.07
A;Map position: 1
A;Introns: 84/2


  Query Match             10.9%;  Score 475.5;  DB 2;  Length 780;
  Best Local Similarity   23.9%;  Pred. No. 6.5e-22;
  Matches  192;  Conservative  149;  Mismatches  299;  Indels  163;  Gaps   36;

Qy      40 ERSGDQVLPFLLQCAEQ----LPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALD 95
           | :  : :  :| | :     :| |||   |||   :|     |  ::|   | | : 
Db      67 EANDPETISTILDCIQTTAFIIPVKIPHLATLIIRASLR----VPLILEKAAAYFCLQYF 122

Qy      96 SGNCNSI-----RILLRFMTSLLC-SKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQP 149
           : | ||      :: ||    |:| |  :|| :|    : | |  |  : :|      | 
Db     123 T-NLNSFLYYEAKVDLRM---LICMSFALQPGTL----KPLFSLLADAISKETKPSVW-- 172

Qy     150 QADFYVICILSSLPW--------GGSELAEQVPDEIERVLVGIQAYLSIRKNSST--SGL 199
             | ::  || :||:         |  : | :: |: |         |:  ||:| |  : | 
Db     173 -GDNFLRIILINLPYFIAANNDLGKKDFANEILDQCE-------IYVRHRKSSITLSNPL 224

Qy     200 NFFHNGEFESSLAEKDFVEDLLDRIQSLA-SNGWKLESVPRPHLSFEAQLVAGKFHELRP 258
           :  |      |:|::   ||| :   |:  | :    : :|   ||:  |    | : 
Db     225 SIHDN------LSEEEL--DLLYKQLILSRENDFTFPYISQPWKFFESDFV----HIV-- 270

Qy     259 IKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLL 318
                 |  ||      :  : |  : :     | |  '     |  :   :  : 
Db     271 -----PVSPSIPEWTFQPTPQQNELPSFKRFFELFNNFEIRTTPDASDVAASIFRDISV 324

Qy     319 DVLFYLNGCRKECASYMANLPVTFRY------------------------EYLMAETLF 353
           ||: :|   | | |  : :| | | |                          | :: | : 
Db     325 DVINHLEFNRVEAAQVLTDLDVYFTYKTFALRGTPVNELPNLDPSESRWKAEDIIVEAVL 384

Qy     354 SQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILW 413
           ::|     :| :||  :::: |:  |          :| ::   ||| :::  | | | 
Db     385 GELLGSQNTTYKPVYYHSLLIECCRIAPKILAPTFGRVIRLMYTMSSDLPLQTLDRFIDW 444

Qy     414 FSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALE 473
           |||||||| | | | ||   ::|    ||:||::| : ||: |||: :|  |:      | 
Db     445 FSHHLSNFNFHWKWNEWIPDVELDDLHPKKVFMRETITRELILSYYTRISDSLP-----E 499

Qy     474 ELL----PPKAGPNFMYSLEEGKEKTEEQQL--SAELSRKVKEKQTARDMIVWIEETIYP 527
           ||        :||||:|  |       :  |:  :  | : ::|     |:|:  || 
Db     500 ELRCLLGEQPSGPNFVYENETHPLYQQSSQIIEALRLHKPLEE----LDIILQSEE---- 551

Qy     528 VHGFEVT-LTIVVQTLLDIGSKSFTHLVTVLERYGQV---FSKLCPDNDKQVMLLSQVST 583
           :   | : : :|:     :||:|||:| : | |::      ||:     | :: :: :: : 
Db     552 IQNSETSAVRLVMSCAYSLGSRSFSHALNVFEKHLNTLKHFSR--KSLDSEIEVVDELFS 609

Qy     584 YWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNR 643
           :||       : :|:|: | ::|  :|: |:   ::|  :  |   |  :  :|:|: 
Db     610 FWKLQPFNAVMWLDKMLNYSIISITSIIEWLIK-QDVTIWSRS-YTWSL----VNTTFNK 663

Qy     644 I-SDLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTV 702
           : : ||: :||    : |: |      |  | |   : |                 : |:: : 
Db     664 LAARLRRSVSNKEDSSLINE-----ANEEKEIVTNLL--------------LSALRALI 703

Qy     703 EKTGEAELSLRESLEAKEALLNRALSETEVLLLLLFQS-FLGVLKERLPDPTKVRSVQDL 761
           :  |                 |  :|    |:|   :| || | |: : :  :   ||: 
Db     704 SENAE----------------NIWVSHWLNLMLKYVESNFLSVKKDTIEEANE--PVQEN 745

Qy     762 KSIGAEDDKPSAMD-VD---SEN 780
           |   || |   :| ||   |||
Db     746 TSEEQEDTKMQPVDAVDEQPSEN 768


RESULT   5
A44919
GCR3 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YM8564.07; protein YM9553.01; protein YMR125w
C;Species: Saccharomyces cerevisiae
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 29-Oct-1999
C;Accession: A44919; S53055; S54494
```

```
R;Uemura, H.; Jigami, Y.
J. Bacteriol. 174, 5526-5532, 1992
A;Title: GCR3 encodes an acidic protein that is required for expression of glycolytic ge
A;Reference number: A44919; MUID:92380925
A;Accession: A44919
A;Molecule type: DNA
A;Residues: 1-858 <UEM>
A;Cross-references: GB:D10224; NID:g464221; PIDN:BAA01076.1; PID:d1001545; PID:g464222
A;Note: sequence extracted from NCBI backbone (NCBIN:112104, NCBIP:112106)
R;Badcock, K.; Churcher, C.
submitted to the EMBL Data Library, March 1995
A;Reference number: S53055
A;Accession: S53055
A;Molecule type: DNA
A;Residues: 339-858 <BAD>
A;Cross-references: EMBL:Z48622; NID:g728663; PIDN:CAA88550.1; PID:g728664; MIPS:YMR125w
R;Lye, G.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A;Reference number: S54014
A;Accession: S54494
A;Molecule type: DNA
A;Residues: 'MFNRKRRG',6-489 <LYE>
A;Cross-references: EMBL:Z49273; NID:g809577; PIDN:CAA89274.1; PID:g809584; MIPS:YMR125w
C;Genetics:
A;Gene: SGD:STO1; GCR3
A;Cross-references: MIPS:YMR125w; SGD:S0004732
A;Map position: 13R
C;Keywords: DNA binding; nucleus

  Query Match             6.4%;  Score 279;  DB 2;  Length 858;
  Best Local Similarity   21.4%;  Pred. No. 1.6e-09;
  Matches  156;  Conservative  117;  Mismatches  242;  Indels  214;  Gaps   34;

Qy      95 DSGNCNSIRILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFY 154
           ::|  | |:::|||:: |  | :     || ::::|   :    : : ||    | ::
Db     137 ETGPWNKIKLILRFLSIL--SPMFLVDELINIYKSLFELSIELNNLDPGNR--VPLSEAI 192

Qy     155 VICILSSLPW------GGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFF--HNGE 206
           | ::|:       |  :| :       : ||:      |: :|    :|||
Db     193 YTNTLLNIPYLFFFNRNNDGLRTKVEE--------LLAYVEQNYLVKTTDINLLREYNGE 244

Qy     207 FESSLAE----------KDFVEDL-----------------------------LDRIQ 225
           : |          |  : :|                               :| ::
Db     245 PPYEMVELVRVVLPNVKKALINNLEQLNELFPDWNHLLTPQTGDEGFNDALTLPSVDDLK 304

Qy     226 S---LASNGWKLESV-PRPHLSFEAQL--VAGKFHELRPIKCMEQPSPPSDHSRAYSGKQ 279
           |   | |   ::|:   |  :|   |   || |  : ||              |:|:
Db     305 SFVRLNKNFGSVDSMWKTPRYAFHVYLPNSAGNFETVVPIS------------TYAGQL 351

Qy     280 KHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCRKE--------C 331
           :| :      |  :     |: |       |  : :   ||:  |   |
Db     352 FNDII------IDLVESLEFNRKE--------VARQVITLDLFFKAGIFTEPGESIAQLI 397

Qy     332 ASYMAN-LPVTFRYEYLMAETLFSQILLLP--QPPFKTLYYTLVIMDLCKALPGAFPAVV 388
           |:|  | |  ||: | |  ||:   |  ||    ||   :|||:: |:|:  | |   |
Db     398 ATYEENPLAPTFKIEDLAIETILGLIFKLPSVSQPF-AYFYTLLV-DICQNSPKAIAPVF 455

Qy     389 AGAVRALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWA-----PKR 443
           | |.  :  :  || | : | : |||  :||| | | | ||    |  |:      ||
Db     456 GRAFRFFYSHLDSLDFELKLRYLDWFSIQMSNFNFSWKWNEWED--DSIKFGKYFYNPKV 513

Qy     444 VFVQEILQREVRLSYWDKIKQSIENATALEELLPPK---------------------- 479
           | : ::|:|:||:          | : :|: || :
Db     514 NFAKNLIQKELRLT---------SNFSEVEDSLPQEFTKYLDTSYIPRDQLINYYQSLFT 564

Qy     480 --------AGPNFMYSLEEG--KEKTEEQQL----SAELSRKVKEKQTARDMIVWIEETI 525
                  | :|  :||   | |  : |     | ||:| | ::       |   :
Db     565 GYTVEEDSVRKNDLYFRQEGVPMENTVRKILDYTHKANNSREVTELES-------ILGEL 617

Qy     526 YPVHGFEVT------LTIVVQTLLDIGSKSFTH----LVTVLERYGQVFSKLCPD-NDKQ 574
             :|  ::      : ::|| : | ||:| :|    :  : |    :|:|:  |   |:
Db     618 KNEYGSIISDFNRFVIILLVQAVTDSGSRSLSHANKYINDLKEDLKTIFAKIELDIETKE 677

Qy     575 VMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSP---------------- 617
           :::  | |:|  | |   :  |      |:::: |  ::|:
Db     678 YIIIEAVLTFWNANPQTGFLVADAFKYAGLLTSRTIFTFIFNETGLKNNGLIEATAIEAV 737

Qy     618 -ENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAE 676
            |: |  :|::     ||     : |:      |:| | ::|       ||  ::|
Db     738 FRNLSQ-QISEE--NESGNNFEFVFERLC----TIANSTIDLL-----DVNADEDIE--- 782

Qy     677 SKLSLVEGE 685
           :  | ||
Db     783 --IPKVNGE 789


RESULT   6
C71083
conserved hypothetical protein PH0929 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 02-Feb-2001
C;Accession: C71083
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; S
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Og
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A;Reference number: A71000; MUID:98344137
A;Accession: C71083
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-879 <KAW>
A;Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA30025.1; PID:g3257342
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenE
C;Genetics:
A;Gene: PH0929
C;Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

  Query Match             3.8%;  Score 166;  DB 2;  Length 879;
  Best Local Similarity   19.3%;  Pred. No. 0.021;
  Matches  106;  Conservative  105;  Mismatches  191;  Indels  146;  Gaps   25;

Qy     213 EKDFVEDLLDRIQSLASNGWKLESVPR-----PHLSFEAQLVAGKFHELRPIKCMEQPSP 267
           || | | :|: |::::||      :||      :  | : :   |: :  :|
Db     202 EKSFTE-VLNEIRNISSN------LPRLRRELEGIKEEVKTLEATFNSITELKL------ 248

Qy     268 PSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGC 327
                 :||:      |  :|||:|      | :  : ::  | |          |
Db     249 ---RLGELNGKK-----GRLEERIRQLESGIEEKRKKSKELEEVVKE---------LPEL 291

Qy     328 RKECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAV 387
           |:   |      : |: |||: :                          :| | |      :
Db     292 EKKETEY--RRLIEFKDEYLVKKN----------------------ELEKRL-----GI 321

Qy     388 VAGAVRALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQ 447
           ::  :: :  ||   | ||:   | |    |   |                       :
Db     322 LSNRLQEVKRKIK--DAESKVARIRWIEERLKEIQ----------------------E 355

Qy     448 EILQREVRLSYWD---KIKQSIEN-ATALEELLPPKAGPNFMYSLEEGKEKTEEQ----- 498
           :|:: | |:  :: .  ::|  :|:  : |  | | |     :| ||  |:: ||:
Db     356 KIMKLEPRVREFEDAMRLKAQMESLKSKLGGLEPEKINEKLLY-LENRKKELEEEIDKIT 414

Qy     499 QLSAELSRKVKEKQTARDMIVWIEET--IYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTV 556
           :   ||::: |::: |   |: :::     || | |:|                |   :
Db     415 RKIGELNQRSKDRRLA---IIELKKARGKCPVCGRELT--------------EEHKADL 456

Qy     557 LERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAI-VRWVF 615
           | :|    | :    :|::     :   :   :     : |:   : :::| | :|
Db     457 LRKYSLELSSI----EKEIQEAKALERQLRAEFRKVENELSRLSSLKTIADQIIEIRERL 512

Qy     616 SPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNIT--KNVLVAEKASANARVELE 673
           |  |::      : :|:| :  ||     :  |:|::: :   ||       |    :|::
Db     513 SKINLEDLKRDKEEYELLKSESNKLKGEVESLKKEVNELNDYKN------ESTKLEIEID 566
```

```
Qy    674 AAESKLSLVE------GEPVLGENPAKMKRL-----KSTVEKTGEAEL-SLRESLEAKEA 721
          |: :|| :|      |   : |   ::: |     |    | || : |||: :
Db    567 KAKKELSEIEDRLLRLGFKTIDELSGRIRELEKFHNKYIEAKNAEKELRDILESLKDERE 626

Qy    722 LLNRALSE 729
          |::|  |
Db    627 ELDKAFEE 634


RESULT   7
T26467
hypothetical protein Y11D7A.14 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C;Accession: T26467
R;Steward, C.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z20218
A;Accession: T26467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1413 <WIL>
A;Cross-references: EMBL:AL032632; PIDN:CAA21588.1; GSPDB:GN00022; CESP:Y11D7A.14
A;Experimental source: clone Y11D7A
C;Genetics:
A;Gene: CESP:Y11D7A.14
A;Map position: 4
A;Introns: 17/3; 62/3; 122/2; 178/3; 200/1; 312/1; 351/3; 387/1; 463/3; 561/1; 583/3; 62
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;35-696/Domain: myosin motor domain homology <MMO>


  Query Match             3.7%;  Score 163;  DB 2;  Length 1413;
  Best Local Similarity   18.1%;  Pred. No. 0.064;
  Matches  160;  Conservative  141;  Mismatches  301;  Indels  284;  Gaps   39;

Qy     13 EKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFL-LQCAEQLPHKIPLYGTLIGLL 71
          |:| :: |  :: :|::    :| : :      :| || :|  :|:     |:   ||
Db    440 EEGIKW-TQVNFANHLQPTIDLIEKPM-----GILSFLEEECV--VPN-----GSEKSLL 486

Qy     72 -----NLENED----------------FVQKLVESVHANFQVALDSGNCN---SIRILL 106
               || |:                 :||    || |    ||    |   |:  :|
Db    487 EKLCSNLSNDSSFKKSKQTQKCSTIRHFYVQHYAGEVHYNIDGWLDKNRDNVETSVLDIL 546

Qy    107 RFMTSLLCSKVIQPA----------------------SLIVVFETLLSSAATTVDEEKG 143
             |  |   :  |                       |  : :|| ||:|
Db    547 SQSTHPLLKLLFPPVPVNNLKTRRGTITNSTVSFLYKNQLQCLLDTLNSSSA-------- 598

Qy    144 NPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFFH 203
                :: |::|:         |::| :|:  ||  |    :: |    |:
Db    599 ---------HFIRCVVSNY--------EKLPGKIDAPLVLAQ----LKCNGVLEGIRICR 637

Qy    204 NGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELRPIKCME 263
          | : | |:  :|:|    |   |  |
Db    638 EG-YPSRLSHSEFIE----RYSLLMKN------------------------------ 659

Qy    264 QPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFY 323
               :  |:  | |:|   :                    :| |
Db    660 -----KEQSKGASEKEKCTLI----------------CQDAQ---------------- 680

Qy    324 LNGCRKECASYMANLPVTFRYEYLMAET-LFSQILLLPQPPFKTLYYTLVIMDLCKALPG 382
             |||                | :| || :: :: :   |   |
Db    681 ---VRKE-------------RYAVGKTKLFCKVGVISELETKRNNYI----------S 712

Qy    383 AFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHLSNFQFI--WPWEEWAFVLD--LPK 438
          :|  ::   :| | :   ::   :   ::   :: | :  |||     :    :|:
Db    713 SFIILIQANIRYLNIQKDLIERRKKLEAVVTIQDNVRQFAELSQWPWYRIYHLTRGLIPR 772

Qy    439 WAPKRVFVQEILQREVRLSYWDKIKQ-SIENATALEELLPPKAGPNFMYSLEEGKEKTEE 497
             |   ::|:   :::|   ::|::  |:|  ||.:|        |   |:   :|||:|:
Db    773 NRDKE-RIEELENEKLKLE--EEIQEMEIKNEEALKE--------NLKLSMLLDREKSEK 821

Qy    498 QQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVL 557
```

```
          ::  ||    |: :        |: :     |  |:  : |          :  ||
Db    822 VKVQKELEEVEKQGR---------EKLLEKEREFRKTMEEMEQN---------EEIFNVL 863

Qy    558 ER-YGQVFSKLCPDND------KQVMLLSQVSTYWKNNVQMTAVAIDRM-MGYRLVSNQA 609
          || | :   |:  ||      :::  |:   |  :| |       :|    | :  :
Db    864 ERKYNEQHKKVMKMNDVLREYERKIEQLNMEKTDLENENQKLRETQNRQDSHYSNLEKEV 923

Qy    610 IVRWVFSPENVDQFH-VSDQPWE------ILGNALN-------KTYNRISDLRKDISNIT 655
          : :    |  :|   :||:  |       |  ||      . :  | | |::| || :
Db    924 MEKSSLIDELQNQIQKLSDENNEQRLTIAKLETALEDEKARFARQNNTIGDMQKLISELN 983

Qy    656 KNVLVAEKASANAR-----VELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAEL 710
          : :   :  : | |     :| | :    |   : :: :   |:  ||    |
Db    984 EKIARFDNIALNERNSTRKIEREKEKLNEELTTAKEIIQKQAKKIDELKEECRKRKNEAS 1043

Qy    711 SLRESLEAKEALLNRALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDK 770
           |   || |||::   : | :              ||||  :    : |:|:|   ::
Db   1044 RLERKLEDKEAMMADCVKELK----------DSHKERLKE--MEQKVEDVKRKNSK--- 1087

Qy    771 PSAMDVDSENGNPKKSCEVGERE-----QWCLSTLGYLTAFTRQYA 811
                :::||   |   |  :||      :  |: | |:   |||:
Db   1088 -----LENENSTQKSQIETFQRESSVDSDYGRSSSGRLSTLGRQYS 1128


RESULT   8
AE1643
ATP-dependent dsDNA exonuclease SbcC homolog sbcC [imported] - Listeria innocua (st
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AE1643
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blc
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsil
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wel
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AE1643
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1023 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC96917.1; PID:g16414173; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: sbcC


  Query Match             3.3%;  Score 145;  DB 2;  Length 1023;
  Best Local Similarity   16.8%;  Pred. No. 0.54;
  Matches  155;  Conservative  159;  Mismatches  297;  Indels  312;  Gaps   38;

Qy      6 TLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIER---SGDQVLPFLLQCAEQLPHKIP 62
          ||: |:              || |      : ::|| :   :|    |   :  | :  ::
Db     80 TLVFRL------------KDKIYQISRIPQQEIAKQRGNGTTTSPQKAELYELIGDEMK 126

Qy     63 LYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRF----MTSLLCSKVI 118
          |  : :  :| : |: :|                  | :  | :|         || |
Db    127 LLASSVSDVNTKMEELIQL---------------NVDQFRQILMIPQGEFRELLVSDSK 170

Qy    119 QPASLIVVFETLLSSAATTVDEEK-GNPSWQPQADFYVICI--------LSSLPWGGSEL 169
          :        | :|   | ||  ||   |  |: | :  ::  :          |: |   | |:
Db    171 EK-------EVILQRLAHTVYYEKVENLLWEKQKEAEILVVEARKKVAELAELSLPGIEI 223

Qy    170 AEQVPDEI-----------ERVLVGIQAYLS-IRKNSSTS--------------GLNFF 202
             :   ||           : :|  ::: || ||| :| :                |: :
Db    224 TGKTTTEISLLQTEAIKQEQAILAELESALSIIRKETSEAVEKVTLAKEQLLDWQNLDLY 283

Qy    203 HNGEFESSLAEKDFVEDLLDRIQSL--ASN-----------GWKLESV--PRPHLSFEAQ 247
             |     ||||| : : :||::   |||          :||:       :: | :
Db    284 IE-EVAKLEAEKDFYQVIANRIEAAKRASNLRSQDALCIRLKEQLETAVNTEKQVAHEVE 342

Qy    248 LVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANK-----M 302
```

```
Db    343 LITNQFS--------------------NAKKQKEALAEQEA------VLEANKRTLFQL 375

Qy    303 EDVQP----IDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILL 358
Db    376 EEMEPKIIELETVTIQKRRAEINW------KEAAD---------RLEKVM---------- 410

Qy    359 LPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHL 418
Db    411 ------KT-----------------EQEIAAELQSMETRLGEINQ-------------- 432

Qy    419 SNFQFIWPWEEWAFVLDLPKWAPKRVFV---QEILQREVRLSYWDKIKQSIENATALEEL 475
Db    433 ---------AELAILEAINKRTTAEALIEKNQELVNKRMKMDLWNKQKQTEEQ--TLKQL 481

Qy    476 LPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTL 535
Db    482 LTEKLAIETMIKQE---ESNLQQEQAATLAAHLHEGDAC-----------PVCG----- 521

Qy    536 TIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCPDN--DKQVMLLSQVSTYWKNNVQMTA 593
Db    522 SVSHPQLAEYGESA--SLTTLEEAKAKLHEKQLAINEVEKSISQLEWQLAEWADIADVDL 579

Qy    594 VAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISN 653
Db    580 VAVEKL----LAENRQLAKNLTDQINQLQTNVAQK--ENIQGALETLKNKQKELETEKNN 633

Qy    654 ITKNVLVAEKASANARVELEAAESKLSLVEG---------------------------- 684
Db    634 TALQV-------ESFHQEVQLSSGKLSYLEQAIPADLRDKTIFDKKKNELSNSIKTHLEQ 686

Qy    685 ----EPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLFQS 740
Db    687 AEQVDKIFREAEKETTRLESTLQSAENTTLDAKEALQVQREVFKESMKQND------FPT 740

Qy    741 FLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTL 800
Db    741 Y-EAYKQAL------MSVEELKS--------------------KEEQLADFERKRHLAI 772

Qy    801 GYLTAFTRQYASEIWPHMEKLES 823
Db    773 SRQADLTEKLKNKQKPNIEQLES 795
```

RESULT 9
A56539
giantin - human
N;Alternate names: macrogolgin
C;Species: Homo sapiens (man)
C;Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999
C;Accession: A56539; S37536
R;Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
Mol. Cell. Biol. 14, 2564-2576, 1994
A;Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein
A;Reference number: A56539; MUID:94187728
A;Accession: A56539
A;Molecule type: mRNA
A;Residues: 1-3259 <SEE>
A;Cross-references: EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g405715
C;Genetics:
A;Gene: GDB:GOLGB1; GCP; GCP371
A;Cross-references: GDB:454958
A;Map position: 3q13.31-3q13.31
C;Superfamily: giantin
C;Keywords: coiled coil; Golgi apparatus; transmembrane protein
F;3238-3254/Domain: transmembrane #status predicted <TMN>

Query Match 3.2%; Score 139.5; DB 1; Length 3259;
Best Local Similarity 19.3%; Pred. No. 6.4;
Matches 168; Conservative 148; Mismatches 356; Indels 197; Gaps 32;

```
Qy      8 LLRIGEKGPEYGTSSDYKDHIETC--------------------FGVIRREIERSGDQV 46
Db   1461 LCEMKQKPEEIGEESRAKQQIQRKLQAALISRKEALKENKSLQEELSLARGTIERLTKSL 1520

Qy     47 LPFLLQCAEQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILL 106
Db   1521 ADVESQVSAQNKEKDTVLGRL-ALLQEERDKLITEMDRSLLENQSL---SSSCESLKLAL 1576

Qy    107 RFMTSLLCSKVIQPASLIVVFETLLSS--AATTVDEEKGNPSWQPQADFYVICILSSLPW 164
Db   1577 EGLTE-------DKEKLVKEIESLKSSKIAESTEWQEKHK---ELQKEYEILL------- 1619

Qy    165 GGSELAEQVPDEIERVLVGIQA--------YLSIRKNSSTSGLNFFHNGEFESSLAE-KD 215
Db   1620 ---QSYENVSNEAERIQHVVEAVRQEKQELYGKLRSTEA-------NKKETEKQLQEAEQ 1669

Qy    216 FVEDLLDRIQSLA-SNGWKLESVPRPHLSFEAQLVAGKFHELRPI-----KCMEQ--PSP 267
Db   1670 EMEEMKEKMRKFAKSKQQKILELEEENDRLRA--------EVHPAGDTAKECMETLLSSN 1721

Qy    268 PSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGC 327
Db   1722 ASMKEELERVKMEYETLSK---KFQSLMSEKDSLSEEVQDLKHQI---------------- 1763

Qy    328 RKECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAV 387
Db   1764 -EDNVSKQANLEATEKHD----------------------NQTNVTEEGTQSIPGE---- 1796

Qy    388 VAGAVRALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQ 447
Db   1797 --------TEEQDSLSMSTRPTC----SESVPSAKSANP----AVSKDFSSHDEINNYLQ 1840

Qy    448 EILQREVRLSYWDKIK-------QSIEN-ATALEELLPPKAGPNFMYSLEEGKEKTEEQQ 499
Db   1841 QIDQLKERIAGLEEEKQKNKEFSQTLENEKNTLLSQISTKDGELKMLQEEVTKMNLLNQQ 1900

Qy    500 LSAELSRKVKEKQTARDMIVWIEETIY----PVHGF------EVTLTIVVQTLLDIGSKS 549
Db   1901 IQEELSRVTKLKETAEEEKDDLEERLMNQLAELNGSIGNYCQDVTDAQIKNELLESEMKN 1960

Qy    550 FTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQA 609
Db   1961 LKKCVSELEEEKQQLVK------EKTKVESEIRKEYLEKIQ----GAQKEPGNKSHAKEL 2010

Qy    610 IVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRI----SDLRKDISNITKNVLVAEKAS 665
Db   2011 QELLKEKQQEVKQLQKDCIRYQEKISALERTVKALEFVQTESQKDLEITKENLAQAVEHR 2070

Qy    666 ANARVELEAAESKLSLVEGEP--VLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALL 723
Db   2071 KKAQAELASFKVLLDDTQSEAARVLADNLKLKKELQSNKE-------SVKSQMKQKDEDL 2123

Qy    724 NRALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNP 783
Db   2124 ERRLEQAEEKHLKEKKN----MQEKL-DALRREKVHLEETIG----------EIQVTLN 2167

Qy    784 KKSCEVGEREQWCLSTLGYLTAFTRQYAS 812
Db   2168 KKDKEVQQLQENLDSTVTQLAAFTKSMSS 2196
```

RESULT 10
T30335
KLP2 protein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 02-Jun-2000
C;Accession: T30335
R;Boleti, H.; Karsenti, E.; Vernos, I.
Cell 84, 49-59, 1996
A;Title: Xklp2, a new Xenopus centrosomal kinesin-like protein required for centrosc
A;Reference number: Z20827; MUID:96140639
A;Accession: T30335
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA

```
A;Residues: 1-1388 <BOL>
A;Cross-references: EMBL:X94082; NID:g1129172; PID:e213754; PIDN:CAA63826.1
C;Genetics:
A;Gene: klp2
C;Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

  Query Match             3.2%;  Score 139;  DB 2;  Length 1388;
  Best Local Similarity   20.0%;  Pred. No. 2;
  Matches  174;  Conservative  156;  Mismatches  328;  Indels  214;  Gaps   45;

Qy      35 IRREIERSGDQVLPFLLQCAEQLP-----HKIPLYGTLIGLLNLENEDFVQKLVESVHAN 89
           :: |::: :|:   |   : |:|      ::|  |         | |::   :|::
Db     381 LQAEVKKLKEQLSQLL---SGQMPGDISVARVPSVGD-------NNMDYMNNFIEAM--- 427

Qy      90 FQVALDSGNCNSIRILLRFMTSL--LCS---KVIQPASLIVVF-ETLLSSAATTVDEEKG 143
           : |:  :    ::||: :  |  ||:   | ||   :|| | |  :|       | :
Db     428 --MILEKSD-REKKVLLQKVVQLEDLCNKKEKFIQSNKMIVKFREDHISRLEKAHKEGRI 484

Qy     144 NPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFFH 203
           : |   | ||     :: |      | |||         :: :    :|     |
Db     485 SLSNNEQDDF-----IAELKEEIRTLKEQVEHHPRVAKYALENHSLREENKRLHSLQSVK 539

Qy     204 NGE-------------FESSLAEKDFVEDLLDRIQSLASNGWKLES--------VPRPH 241
           :               | |::|||       ::  : |   :|::        :
Db     540 RAQEVTAQMIAELEKAFLEVSVSEKD------RQVAPMHSTPIQLDNNSLMSAARMRERM 593

Qy     242 LSFEAQLVAGK-----FHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRR-LN 295
           |  |::|   |     | ||   | :|| |           :  :| :  | :   |
Db     594 LQLESELATSKQEYEEFKELTKKKQVEQES------------ELQSLIKSNQHLENILE 640

Qy     296 IFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFR----------YE 345
           |||  :|  ::|    | : ::           ||     :  |           :
Db     641 AIKANKRHEVSQLNRMHAAETIKNM-------TTPTKSYNLRSRLVPRLSPDAMPNGLMD 693

Qy     346 YLMAETLFSQILLLPQPP------FKTLYYTL-VIMDLCKALPGAFPAVVAGAVRALFEK 398
           :  :   |:  | ||      :: :   | :: :  ||           :| | ::
Db     694 TPKSGDVMDDIINEPIPPEMSEQAYEAIAEELRIVQEQVTALQAKLDEEEGKNIR-LQQQ 752

Qy     399 ISDLDMESRTRLILWFSHHLSNFQFIWPWEEW---AFVLDLPKWAPKRVFVQEILQREVR 455
           :: |:: | |::   |:   ||    |  |:     | :  | |   :    :::|: ||
Db     753 VNKLELCS-TQIQELFNSERSN----WNKEQQDLIAQIKSLEKQKQENKSQEDVLKSEVH 807

Qy     456 LSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSA------------ 502
           ::  :::|    | ||  |      |||   |:: |  ||||
Db     808 -----DLRVVLQSAD--RELGAVKG----EYSLYREKQEKELSQLSARHMDVQLQLDNVR 856

Qy     503 -ELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSK--------SFTHL 553
            |    ::||:: :|    :||    |  ||:    : | : |  ||          |::|
Db     857 LEHETLLEEKRSLQDAFDNLEE----VMKFEIDQ--LKQEISD--SKHENETLRAEFSNL 908

Qy     554 VTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRW 613
           : :||           : :::  | ||:    :|  :      :|  |  |   :: : :
Db     909 LELLE----------TEKERRQKLTSQLEEDKENKTKELLQVVDENMHLRKQCSELMTK- 957

Qy     614 VFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRK----------DISN---ITKNVLV 660
           : | : |        | ::|  :   |:|| |           |: |   : :  ::
Db     958 --CEQQVTELHG-------LEHSLTSSKEMIADLEKKNTADKEVVADLMNQIQVHRTTII 1008

Qy     661 AEKASANARV-ELEAAESKLSLV-----EGEPVLGENPAKMKRLKSTVEKTG-----EAE 709
           :  | :    |||   || |:|     | : |: |   ::: |:  :|:       | |
Db    1009 HKTESIDLLTRELEDIHSKYSIVLLAKEESKTVIEEQEKQIEELRECLERKQSADNIEKE 1068

Qy     710 LSLRESLEAKEAL--LNRALSETEVLLLLLFQSFLGVLKERLPDPT-KVRSVQDLKSIGA 766
           |   :   | | |  | | :: |  ||   :  |   :::: : | :|: : ||: |
Db    1069 LLCDDLAHATEELEKLTEAFNKQEA-LLHTHEKELVEKEQQISELTNQVKLMTDLE-ISR 1126

Qy     767 EDDK--PSAMDVDS----------ENGNPKKS 786
           | :|  |:: :  |            |||  |
Db    1127 EQEKIRPASSNSSSPVVLPETPRTPEGNPYDS 1158

RESULT  11
```

```
T24587
hypothetical protein T06E4.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24587
R;Lloyd, C.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z19910
A;Accession: T24587
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1295 <WIL>
A;Cross-references: EMBL:Z70756; PIDN:CAA94789.1; GSPDB:GN00023; CESP:T06E4.1
A;Experimental source: clone T06E4
C;Genetics:
A;Gene: CESP:T06E4.1
A;Map position: 5
A;Introns: 17/3; 218/2; 385/3; 481/3; 946/3; 1034/3; 1222/2

  Query Match             3.2%;  Score 138;  DB 2;  Length 1295;
  Best Local Similarity   19.9%;  Pred. No. 2.1;
  Matches  173;  Conservative  145;  Mismatches  307;  Indels  246;  Gaps   42;

Qy      20 TSSDYKDHIETC---FGVIRREIERSGDQVLPFLLQCAEQLPHKIPLYG-----TLIGL- 70
           ||   |  ||     | :: : ||   |:|:       : |  || | |     ||: :
Db     206 TSQHEKSRIEYSLYEFEMLEKNIEH--DEVV-------KDLQLKIELLGEKHRQTLVEIK 256

Qy      71 --LNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLIVVFE 128
             :| :  :|:: |:|    :   :     ::  ||       |  |: :|    : ||
Db     257 EAREVEEKMLMQQVEEAVK---KTKSEREAAKTLEQLLEKRIEELEMKLTEPNGEKLQFE 313

Qy     129 TLLSSAATTVDE-----------------EKGNPSWQPQADFYVICILSSLPWGGSELA 170
           |    :  :|                   ||     |                    |:
Db     314 HQLEELKSRCEELTDKALKVDVMQHSIEDYEKKFVELQ-------------------EMK 354

Qy     171 EQVPDEIERVLVGIQA----YLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDL-LDRIQ 225
           |:  :::::    |:     |: :   :|     | | | |:  :| : |  | || |
Db     355 EEADEQLQKAKEDIETLQMKYVEL---ETTINKEVFSNSEIETLKSEHEIVRKLMLDEIH 411

Qy     226 SLASNGWKLESVPRPHLSFEAQLVAGKFHELRPIKCMEQPSPPSD-HSRAYSGKQK---- 280
           | :   :: ::   : : | : :      :|: | |     |   : |     |:|
Db     412 RLEN---EMSALQPKNDTTELEELQKTLDDLK-IDCCNLTSKMLELQSELVEVKEKATSE 467

Qy     281 -HDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCRKECAS---YMA 336
            :|: :  : : ::|   : ::|: :          |:|:   ||   ::       ::
Db     468 IGEAVQKNGELLEQIN---SLRVENAK----------LVDMEGQLNDAHRKAEDKDVRIS 514

Qy     337 NLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALF 396
           |  |   | |  ::  |  ||:     :  |:| : :    |          :   :|
Db     515 ELLTTI--ESLRQDSEASDKLLMDSESTQN-EYSLALENTVSEL-----ETMRREYKASV 566

Qy     397 EKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRL 456
           :|:  | :|          |  |                          |: : |:|:
Db     567 DKVCSLQLELEE-----IQHETS-------------------------VELEEAEIRI 594

Qy     457 SYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARD 516
                 |   | || :  | | : :  ||  :|  :      :|  ::|||:|
Db     595 K---------ELELAQEEAV--KTGSSQLKKLEIVQEDCQ------KLRDQLKEEQI--Q 635

Qy     517 MIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVM 576
           :| : ||   :|                :|  |     |:| |: |||     : :
Db     636 QLVSLRETSEVMH----------------QESARHQE---EKY-QIQSKLMSTEAEVIE 674

Qy     577 LLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHV----------S 626
           | | :                       | || || | |    || |:          :
Db     675 LRSSID----------------------SLQAEVR-VQSDSAADQKHILEDYLRKIRQA 710

Qy     627 DQPWEILGNALNKTYNRISDLRKD----ISNITKNVLVAEKASANARVELEAAESKLSLV 682
           ::  | | : |  :  :| ||:      | :: : :  ||  :   :| ||  : ::|
Db     711 EETNEKLRSDLASSEEQILDLKNQQESLIDDLKEKLHSAESTNQELQVSLEMLKIEVS-- 768

Qy     683 EGEPVLGENPAKMKRLKSTVEKTG-EA-ELSLRESLEAKEALLNRALSETEVLLLLLFQS 740
```

```
Db    769 ---------NARQKVMESEVLKESFEALQLELSASQEVSRSVVDAAVQEKDGLLRL---- 815

Qy    741 FLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGE--REQWCLS 798
Db    816 -VDTLKLKIEDTEK--SAQDLQQSSVEEIKQLQLDLQ----NFKQNAEVLESLNEKLNSS 868

Qy    799 TLGYLTAFTRQYASEIWPHMEKLESEVFSGE 829
Db    869 HKRDMVALASQ--------LEELQHKLVVGE 891


RESULT  12
I52300
giantin - human
N;Alternate names: gcp372
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C;Accession: I52300
R;Sohda, M.; Misumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A;Title: Molecular cloning and sequence analysis of a human 372-kDA protein localized in
A;Reference number: I52300; MUID:95100974
A;Accession: I52300
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3225 <RES>
A;Cross-references: GB:D25542; NID:g662389; PIDN:BAA05025.1; PID:g808869
C;Superfamily: giantin


  Query Match             3.1%;  Score 137.5;  DB 2;  Length 3225;
  Best Local Similarity   19.3%;  Pred. No. 8.4;
  Matches  168;  Conservative  147;  Mismatches  357;  Indels  197;  Gaps   32;

Qy      8 LLRIGEKGPEYGTSSDYKDHIETC--------------------FGVIRREIERSGDQV 46
Db   1427 LCEMKQKPEEIGEESRAKQQIQRKLQAALISRKEALKENKSLQEELSLARGTIERLTKSL 1486

Qy     47 LPFLLQCAEQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILL 106
Db   1487 ADVESQVSAQNKEKDTVLGRL-ALLQEERDKLITEMDRSLLENQSL---SSSCESLKLAL 1542

Qy    107 RFMTSLLCSKVIQPASLIVVFETLLSS--AATTVDEEKGNPSWQPQADFYVICILSSLPW 164
Db   1543 EGLTE-------DKEKLVKEIESLKSSKIAESTEWQEKHK---ELQKEYEILL------- 1585

Qy    165 GGSELAEQVPDEIERVLVGIQA--------YLSIRKNSSTSGLNFFHNGEFESSLAE-KD 215
Db   1586 ---QSYENVSNEAERIQHVVEAVRQEKQELYGKLRSTEA-------NKKETEKQLQEAEQ 1635

Qy    216 FVEDLLDRIQSLA-SNGWKLESVPRPHLSFEAQLVAGKFHELRPI-----KCMEQ--PSP 267
Db   1636 EMEEMKEKMRKFAKSKQQKILELEEENDRLRA--------EVHPAGDTAKECMETLLSSN 1687

Qy    268 PSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGC 327
Db   1688 ASMKEELERVKMEYETLSKKFQSLMS--------------EKDSLSEEVQDLKHQIEG- 1731

Qy    328 RKECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAV 387
Db   1732 ---NVSKQANLEATEKHD--------------------NQTNVTEEGTQSIPGE---- 1762

Qy    388 VAGAVRALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQ 447
Db   1763 --------TEEQDSLSMSTRPTC----SESVPSAKSANP----AVSKDFSSHDEINNYLQ 1806

Qy    448 EILQREVRLSYWDKIK-------QSIEN-ATALEELLPPKAGPNFMYSLEEGKEKTEEQQ 499
Db   1807 QIDQLKERIAGLEEEKQKNKEFSQTLENEKNTLLSQISTKDGELKMLQEEVTKMNLLNQQ 1866

Qy    500 LSAELSRKVKEKQTARDMIVWIEETIY----PVHGF------EVTLTIVVQTLLDIGSKS 549
Db   1867 IQEELSRVTKLKETAEEEKDDLEERLMNQLAELNGSIGNYCQDVTDAQIKNELLESEMKN 1926

Qy    550 FTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQA 609
Db   1927 LKKCVSELEEEKQQLVK------EKTKVESEIRKEYLEKIQ----GAQKEPGNKSHAKEL 1976

Qy    610 IVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRI----SDLRKDISNITKNVLVAEKAS 665
Db   1977 QELLKEKQQEVKQLQKDCIRYQEKISALERTVKALEFVQTESQKDLEITKENLAQAVEHR 2036

Qy    666 ANARVELEAAESKLSLVEGEP--VLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALL 723
Db   2037 KKAQAELASFKVLLDDTQSEAARVLADNLKLKKELQSNKE-------SVKSQMKQKDEDL 2089

Qy    724 NRALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNP 783
Db   2090 ERRLEQAEEKHLKEKKN----MQEKL-DALRREKVHLEETIG----------EIQVTLN 2133

Qy    784 KKSCEVGEREQWCLSTLGYLTAFTRQYAS 812
Db   2134 KKDKEVQQLQENLDSTVTQLAAFTKSMSS 2162


RESULT  13
B70356
chromosome assembly protein homolog - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Jun-2000
C;Accession: B70356
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666
A;Accession: B70356
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1156 <AQF>
A;Cross-references: GB:AE000699; NID:g2983238; PIDN:AAC06839.1; PID:g2983243; GB:AE0
A;Experimental source: strain VF5
C;Genetics:
A;Gene: xcpC
C;Superfamily: chromosome segregation protein SMC1


  Query Match             3.1%;  Score 137;  DB 2;  Length 1156;
  Best Local Similarity   21.0%;  Pred. No. 2;
  Matches  154;  Conservative  105;  Mismatches  244;  Indels  232;  Gaps   32;

Qy    218 EDLLDRIQSLASNGWKLESVPRPHLSFEAQ--------LVAGKFHELRPIKCMEQPSPPS 269
Db    391 EKLTEKLNSLNKEKQELE-IQRANLKNKIERIKEDINKLISEREEKIKEIKEKEQ----- 444

Qy    270 DHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCRK 329
Db    445 EIKRLKAIKKKEEEELR--NLTQELNIYEKRLSE---------VRKKLEEVLKEKGAIER 493

Qy    330 ECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAV-- 387
Db    494 EVRS------------------FSD-----------------VSDVFKDIKGVYGSVSE 517

Qy    388 --------------VAGAVRALFEKISDLDMESRTRLILWFSHHLSNFQFI--------- 424
Db    518 LIRVKNPEHITAIEVAGGGRLKFIVVEDEEV-AKECIQLAKRMNLGRFSFIPLNRVRVEE 576

Qy    425 ----WPWEEWA--FVLDLPKWAPKRVFVQEILQREVRLSYWDK-IKQSIENATA------ 471
Db    577 RPLRYPRTKGAVDFAVNLVEYDPK-------FEKVVKFVFGDTLIVENFESAKAIGIGNY 629

Qy    472 ----LEELLPPKAGP------------NFMYSLEE-GKEKTEEQQLSAELSRKVKEKQTA 514
Db    630 RMVTLEGELFEKSGVITGGAVKPSGELNKRYYEEELQRLNAEEEKLKNEESIIQKKIREI 689
```

```
Qy     515 RDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTV-LERYGQVF--------- 564
           |::|   |:|                 || :  :    | :  ||:| : |
Db     690 RNLI--SEKT---------------ALLKVSERKIEELSSEGLEQYEEKFKEKLENSKE 731

Qy     565 ------SKLCPDNDKQVMLLSQVSTYWK--NNVQMTAVAIDRMMGYRLVSNQAIVRWVFS 616
                 ||    ||   |  ::  | :  ||:::    | |                :|
Db     732 YLKILEEKLLNVEDKLKELAEEIEYYEEKLNNLKLKEGDIKRH---------------YS 776

Qy     617 PENVDQ-----FHVSDQPWEI------LGNALN-KTY--------------------NRI 644
           | |::       |  |  ||      :   || |||                     ||
Db     777 REGVEEKRREYSKVRKQVSEIEKSLNEIERELNKKTYELEYLEKEIQEKEREREYLTERI 836

Qy     645 SDLRKDISNITKNVLVAEKASANARVELEAAESKL-----------SLVEGEPVLGENP 692
             |:|:| |:   :|  ||       |:: || |:             ::  :  ||:
Db     837 KSLKKEIENL---ILFKEKTLQ----EVKEAEVKVYDYIKQKEELEKEILNLKSKLGKLK 889

Qy     693 AKMKRLKSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLFQSFLGVLKERLPDP 752
           | : ||   ||  | | :|:  || |   ||  | | | | |   :  :  |||:|
Db     890 IKEEELK---EKIFEKEKNLK-VLEEKIENLNEELKEYEDLKLGADEESIPKLKEKLKRV 945

Qy     753 TKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLTAFTRQYAS 812
           |:   :| | |:    ::  | ::   |   :|  :: |  :     :       |:
Db     946 TE--EIQKLGSVNFRAEEDYAEELKRFNDYKEKQQKLKEESKAIKKLIEETENKKRKVFL 1003

Qy     813 EIWPHMEKLESEVFS 827
           | :  : |    :||
Db    1004 EAFNQINKSLKRIFS 1018
```

RESULT 14
T22976
hypothetical protein F59A2.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T22976; T23157
R;Lightning, J.
submitted to the EMBL Data Library, June 1994
A;Reference number: Z19645
A;Accession: T22976
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1133 <WIL>
A;Cross-references: EMBL:Z34801; PIDN:CAA84332.1; GSPDB:GN00021; CESP:F59A2.6
A;Experimental source: clone F59A2
R;Burton, J.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z19700
A;Accession: T23157
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1133 <WI2>
A;Cross-references: EMBL:Z66514; PIDN:CAA91344.1; GSPDB:GN00021; CESP:F59A2.6
A;Experimental source: clone K01A11
C;Genetics:
A;Gene: CESP:F59A2.6
A;Map position: 3
A;Introns: 13/2; 43/3; 107/3; 413/3; 492/3; 567/3; 635/3; 710/3; 738/3; 795/2; 1008/3; 1

Query Match 3.1%; Score 136; DB 2; Length 1133;
Best Local Similarity 22.1%; Pred. No. 2.3;
Matches 102; Conservative 73; Mismatches 182; Indels 104; Gaps 21;

```
Qy     442 KRVFVQEI--LQREVRLSYWDKIKQSIENATALEELLPPK------AGPNFMYSLEEG-- 491
           |   ||::  ||:|: |   | ::  :  | || | :  |              |||
Db      39 KNEMVQQLSRLQQEM-LEKCDALQAEVNEAKALREEIQAKYDDVTQKAERIQGELEESKK 97

Qy     492 -----------KEKTEEQQLSAELSRKVKEKQTARDMIV----WIEETIYPVHGF--EV 533
                     ||:  |:|| |:   |:  :|   | :       || :    |   |:
Db      98 VLESEKQAFENEKEQEREEQL-AKAMEKLNSEQNILDEVTKKLEQSEEEVLAARGAIQEL 156

Qy     534 T--------LTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKL------C-PDNDKQVMLL 578
           |         |   :| |:  ||      | |: :  :   :       |    |: | ||
```

```
Db     157 TEKLEESEKETSTAKTELEAVSKKLDSSETSLKEFSDMIEAMKIQLINCEKQKDEAVELL 216

Qy     579 SQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALN 638
           |     : |                  :|:  : : :        :      :  ||:   |
Db     217 KQKLEEVEKN---------------MSDVEVQKQLLLESTTSEMKQHAEAAEIVKKQLE 260

Qy     639 KTYNRISDLRKDISN--ITKNVLVAEKASANARV--ELEAAESKLSLVEGEPVLGENPAK 694
           :  : | :|:||   |     |  | ::::|| : :  ::|||: :|   | |    |  :
Db     261 EAQSSIENLKKDAENERNLKTALESDESSAISEITKQMEAAKKELEASEKEK--SELREQ 318

Qy     695 MKRLKST----------VEKTGEAELS-LRESLE----AKEALLNRALSETEVLLLLLFQ 739
           | ||:             ::|| | |:: : :| |    |:| |     :  | | ::  :
Db     319 MDRLQKVHNAGQEDIQKLQKTWELEMAKIAKSTEDEKLAREQLAGELENAKEDLKVVEEE 378

Qy     740 SFLG---------------VLKERLP-DPTKVRSVQDLKSIGAEDDKPSAMDVDSENGN 782
             |                |||||:|  . : : | |:| |   : ||   :: : :|
Db     379 KHTGIQRAQGALDDAEKEVKVLKEQLERAQSALESSQELAS-SQKADKIQELEKELQNAQ 437

Qy     783 PKKSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLES 823
           : | |:    :   |    |||      ||     :|||:
Db     438 KRSSEELETANEMVRS----LTATLENSNSETEILKQKLET 474
```

RESULT 15
T13734
groovin gene protein - fruit fly (Drosophila melanogaster) (fragment)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13734
R;Strumpf, D.; Volk, T.
J. Cell Biol. 143, 1259-1270, 1998
A;Title: Kakapo, a novel cytoskeletal-associated protein is essential for the resti
A;Reference number: Z17746; MUID:99054752
A;Accession: T13734
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4151 <STR>
A;Cross-references: EMBL:Y09430; NID:g4185883; PIDN:CAA70581.1; PID:g4185884
C;Genetics:
A;Gene: grv
A;Cross-references: FlyBase:FBgn0013733
A;Map position: 2

Query Match 3.1%; Score 135; DB 2; Length 4151;
Best Local Similarity 19.1%; Pred. No. 17;
Matches 183; Conservative 156; Mismatches 348; Indels 270; Gaps 43;

```
Qy      55 EQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCN--SIRILLRFMTSL 112
           |:|   |  :  |:  ||   |     :  :   |     ||: |    ::|: ||
Db    2482 ERLLQSIAEHKPLLDKLNKTGEALGALVADDDGAKINEILDTDNARYAALRLELRERQQA 2541

Qy     113 LCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQ 172
           | |: : :       | :| : | |||:         | |      ||:||    :: ||
Db    2542 LESALQESSQFSDKLEGMLRALANTVDQVN-------QLD-----PLSALP---QKIREQ 2586

Qy     173 VPDE---------------IERVL------VGIQAYLSIRK-NSSTSGLNFFHNGEFES 209
           : |                 ::|         | :|  ::|   :    ||  |    :
Db    2587 IEDNDALMDDLDKRQDAFSAVQRAANDVIAKAGNKADPAVRDIKAKLEKLNNLWNDVQNA 2646

Qy     210 SLAEKDFVEDLLDRIQSLASNGWK-LESVPRPHLSFEAQLVAGKFHELRPIKCMEQP-SP 267
           :      ::|:|    |:|   || | || :     |  |                | | | :
Db    2647 TKKRGSSLDDIL----SVAEPFWKQLNSVMKTLKDLEETL-----------SCQEPPAAQ 2691

Qy     268 PSDHSRAYSGKQ--KHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDV----- 320
           | |  :     |  :|:     |: : ::    :| |      |:  |:::: |:
Db    2692 PQDIKKQQVALQEIRHEIDQTKPE-VEQVRRHGSNLMNMCGEPDKPEVKKHIEDLDNAWD 2750

Qy     321 -LFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKA 379
            :  |   |:|     ||           ||| : :  | :   |  :   |  |:
Db    2751 NITALYAKREE------NLIDAMEKAMEFHETLQNLLKFLTKAEDKFAHLGAVGSDI--- 2801

Qy     380 LPGAFPAVVAGAVRALFEKISD---------LDMESRTRLILWFSHHLSNFQFIWPWEEW 430
```

```
Db   2802 ----------DAVKRQIEQLKSFKDEVDPHMVEVEALNRQAVELTERTS--------PEQ 2843

Qy    431 AFVLDLP------KW-APKRVFVQEILQREVRLSYWDKIKQSI--------ENATALEEL 475
Db   2844 AASIREPLSVVNRRWEALLRGMVERQKQLEHALLHLGQFQHALNELLVWINKTDSTLDQL 2903

Qy    476 LPPKAGP--------------------NFMYSLEE-GKEKTEEQQLSAELS------RK 507
Db   2904 KPIPGDPQLLEVELAKLKVLANDIQAHQNSVDTLNDAGRQLIETEKGSVEASTTQEKLRK 2963

Qy    508 VKEK-----QTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDI-----GSKSFTHL---- 553
Db   2964 LNNEWKQLLQKASDRQHELEEALREAHGYIAEVQDILGWLGDVDAVIGASKPVGGLPETA 3023

Qy    554 VTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQM--------------------T 592
Db   3024 TEQLERFMEVYNELDENRPKVETIQAQGQEYIKRQNQMKVSSSNLQHTLRTLKQRWDAVV 3083

Qy    593 AVAIDRMMGYRLV---------SNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNR 643
Db   3084 SRASDKKIKLEIALKEATEFHDTLQAFVEWLTQAE-----------KLLSNA--EPVSR 3129

Qy    644 ISD-----------LRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENP 692
Db   3130 VLETIQAQMEEHKVLQKDVSTHREAMLLLDKKGTH----LKYFSQKQDVILIKNLLVSVQ 3185

Qy    693 AKMKRLKS-TVEKTGEAELSLRESLEAKEAL--LNRALSETEVLLLLFQ---------- 739
Db   3186 HRWERVVSKAAERTRALDHGYKEAREFNDAWSGMMQYLQETEQVLDQIIEEATASKEPQK 3245

Qy    740 --SFLGVLKE---------RLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCE 788
Db   3246 IKKYIGKLKETHRQLGAKQSVYDGT-MRTGKNLLERAPKGDRP-VLD--------KMLIE 3295

Qy    789 VGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAISSALQF 845
Db   3296 L--KEQW-----------TRVWSKSIDRQRKLEEALLLSGQ-----FSDALGELLDW 3334
```

Search completed: August 20, 2002, 15:27:37
Job time: 5106 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:58:56 ; Search time 32.69 Seconds
(without alignments)
633.616 Million cell updates/sec

Title: US-09-882-986-2
Perfect score: 4374
Sequence: 1 MSNWKTLLLRIGEKGPEYGT.....EDVHPLFLQAISSALQFPLH 848

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	131.5	3.0	2154	2	US-08-841-349-4	Sequence 4, Appli
2	125	2.9	469	2	US-08-968-751-2	Sequence 2, Appli
3	122	2.8	443	2	US-08-795-475-6	Sequence 6, Appli
4	120	2.7	469	4	US-09-052-089A-1	Sequence 1, Appli
5	120	2.7	893	4	US-09-019-160-2	Sequence 2, Appli
6	119	2.7	1805	1	US-07-853-913-2	Sequence 2, Appli
7	117.5	2.7	893	4	US-09-019-160-9	Sequence 9, Appli
8	116.5	2.7	1786	4	US-08-973-462-8	Sequence 8, Appli
9	116	2.7	708	4	US-09-019-160-5	Sequence 5, Appli
10	116	2.7	893	4	US-09-019-160-6	Sequence 6, Appli
11	116	2.7	893	4	US-09-019-160-7	Sequence 7, Appli
12	116	2.7	893	4	US-09-019-160-8	Sequence 8, Appli
13	113.5	2.6	3248	1	US-08-353-700-1	Sequence 1, Appli
14	113.5	2.6	3248	5	PCT-US95-16216-1	Sequence 1, Appli
15	112	2.6	1333	3	US-09-356-952-2	Sequence 2, Appli
16	111.5	2.5	1312	2	US-08-592-126-148	Sequence 148, App
17	111	2.5	893	2	US-08-706-702-3	Sequence 3, Appli
18	111	2.5	893	3	US-08-706-706-3	Sequence 3, Appli
19	110.5	2.5	2482	1	US-08-328-254-6	Sequence 6, Appli
20	109.5	2.5	677	4	US-09-019-160-3	Sequence 3, Appli
21	109	2.5	893	3	US-08-484-661A-2	Sequence 2, Appli
22	109	2.5	893	3	US-08-656-664-2	Sequence 2, Appli
23	109	2.5	893	5	PCT-US96-09641-2	Sequence 2, Appli
24	108.5	2.5	911	3	US-09-356-952-6	Sequence 6, Appli
25	108.5	2.5	1312	2	US-08-687-080-51	Sequence 51, Appl
26	107.5	2.5	1388	4	US-09-572-191-2	Sequence 2, Appli
27	105.5	2.4	955	1	US-08-006-676B-1	Sequence 1, Appli
28	105.5	2.4	955	1	US-08-282-845-2	Sequence 2, Appli
29	105.5	2.4	955	2	US-08-428-414A-3	Sequence 3, Appli
30	105.5	2.4	955	5	PCT-US94-00324-1	Sequence 1, Appli
31	105.5	2.4	976	4	US-09-104-324B-4	Sequence 4, Appli
32	105.5	2.4	2089	1	US-08-418-893D-23	Sequence 23, Appl
33	105.5	2.4	2089	1	US-08-418-893D-24	Sequence 24, Appl
34	105	2.4	285	1	US-08-149-809-24	Sequence 24, Appl
35	104	2.4	753	4	US-08-026-143B-3	Sequence 3, Appli
36	104	2.4	753	5	PCT-US92-10621-3	Sequence 3, Appli
37	104	2.4	753	5	PCT-US94-02233-3	Sequence 3, Appli
38	103	2.4	753	1	US-08-712-241-3	Sequence 3, Appli
39	103	2.4	1711	3	US-08-369-822C-10	Sequence 10, Appl
40	103	2.4	1711	3	US-08-582-776C-10	Sequence 10, Appl
41	103	2.4	1711	3	US-08-434-831B-10	Sequence 10, Appl
42	103	2.4	2101	1	US-08-466-390-4	Sequence 4, Appli
43	103	2.4	2101	1	US-08-470-950-4	Sequence 4, Appli
44	103	2.4	2101	1	US-08-467-781-4	Sequence 4, Appli
45	103	2.4	2101	1	US-08-195-487-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-841-349-4
; Sequence 4, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
; APPLICANT: MISHRA, LOPA
; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT. . .
; FILE REFERENCE: XX/PO4470US0
; CURRENT APPLICATION NUMBER: US/08/841,349B
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2154
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-841-349-4

Query Match 3.0%; Score 131.5; DB 2; Length 2154;
Best Local Similarity 17.9%; Pred. No. 0.0061;
Matches 179; Conservative 151; Mismatches 283; Indels 389; Gaps 47;

```
Qy    118 IQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEI 177
          |:   :|  :||| |     :::                 | ||::  :  |
Db    285 IETEKMIEKYETLASDLLEWIEQ-------------TIIILNNRKFANS---------- 320

Qy    178 ERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQS-LASNGWKLES 236
             |||:|  |           | :   |      ||  :| ||  ||| : :|  |:
Db    321 ---LVGVQQQL--------QAFNTYRTVEKPPKFTEKGNLEVLLFAIQSKMRANNQKV-Y 368

Qy    237 VPRPHLSFEAQLVA-----------------GKFHELRPIKCMEQPSPPSDHSRAY--- 275
          :||     | :|::                  :||   : :|| :   |   |
Db    369 MPR-----EGKLISDINKAWERLEKAEHERELALRNELIRQEKLEQLARRFDRKAAMRET 423

Qy    276 ---------------------SGKQKHDA----LTRYPQRIRRLNIFPANKME-----D 304
                              :  :||:|    :  | :|:: : :  | ::|     |
Db    424 WLSENQRLVSQDNFGFDLPAVEAATKKHEAIETDIAAYEERVQAV-VAVARELEAENYHD 482

Qy    305 VQPI----DRFV-VEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLMAETLF------ 353
          :: |    |  : : ||||::|      |::       | :    : :  | |:
Db    483 IKRITARKDNVIRLWEYLLELL----RARRQ------RLEMNLGLQKIFQEMLYIMDWMD 532

Qy    354 -SQILLLPQPPFKTLYYTLVIMDLCK--ALPGAFPAVVAGAVRAL-------------- 395
           ::||| |   | |   | : || :  ||  |  |: |  || :
Db    533 EMKVLLLSQDYGKHL---LGVEDLLQKHALVEADIAIQAERVRGVNASAQKFATDGEGYK 589

Qy    396 ----------------FEKISDLDMESRTRL----ILWFSHHLSNFQFIWPWEE---WA 431
                           ::::  |  | | ||     ||        :| |   |   |
Db    590 PCDPQVIRDRVAHMEFCYQELCQLAAERRARLEESRRLW--------KFFWEMAEEEGW- 640
```

```
Qy    432 FVLDLPKWAPKRVFVQEILQREVRLSYWD------KIKQSIENATALEELLPPKAGPNFM 485
                        | ::|  ||  |       : : :  | |: :  ::| :|
Db    641 ----------------IREKEKILSSDDYGKDLTSVMRLLSKHRAFEDEMSGRSG-HFE 682

Qy    486 YSLEEGK----------EKTEE-----------QQLSAELSRKVKE-------KQTARD 516
          :::||:          ||  |           :||||   ::::|       :  | |
Db    683 QAIKEGEDMIAEEHFGSEKIRERIIYIREQWANLEQLSAIRKKRLEEASLLHQFQADADD 742

Qy    517 MIVWIEETIYPVHGFEV------TLTIV----------------VQTLLDIGS------- 547
          :  |: : :  |   :|      | ::|                : || :  |
Db    743 IDAWMLDILKIVSSNDVGHDEYSTQSLVKKHKDVAEEITNCRPTIDTLHEQASALPQAHA 802

Qy    548 ----------------KSFTHLVTVLERYGQ---VFSKLCPDNDKQVMLLSQVSTYWKNN 588
                         |    |  : ::  |      |:  : |   : : :    | ||
Db    803 ESPDVKGRLAGIEERCKEMAELTRLRKQALQDTLALYKMFSEADACELWIDE-KEQWLNN 861

Qy    589 VQM------------------------TAVAIDRMMGYRLVSN------------QAI 610
          :|:                        : ||:   :  :|: |             :
Db    862 MQIPEKLEDLEVIQHRFESLEPEMNNQASRVAVVNQIARQLMHNGHPSEKEIRAQQDKLN 921

Qy    611 VRW------------VFSPENVDQFHVSDQPWEILGNALNKTYNRISDLRK------DI 651
          ||             : |  ::  :|:           |:| : | :  |      |:
Db    922 TRWSQFRELVDRKKDALLSALSIQNYHLE----------CNETKSCIREKTKVIESTQDL 971

Qy    652 SNITKNVLVAEKASANARVELEAAESKLSLV--EGEPVLGENPAKMKRLKSTVEKTGEAE 709
          |    |:  :        :| | |:||| :  | | :  |:| : : : | : :  :
Db    972 GNDLAGVMALQCKLTGMERDLVAIEAKLSDLQKEAEKLESEHPDQAQAILSRLAEISDVW 1031

Qy    710 LSLRESLEAKEALLNRALSETEVLLLLL--FQSFLGVLKERLPDPTKVRSVQDLKSIGAE 767
            :: :|: :|| |  | |: :  |  |  |||:|   :  :                  |
Db   1032 EEMKTTLKNREASLGEA-SKLQQFLRDLDDFQSWLSRTQTAI----------------AS 1074

Qy    768 DDKPSAM--------------DVDSENGNPKKSCEVGEREQWCLSTLGYLTA---FTRQ 809
          :| |: :                ::|:   : :|  ::||     : | |   |   | ||
Db   1075 EDMPNTLTEAEKLLTQHENIKNEIDNYEEDYQKMRDMGE-----MVTQGQTDAQYMFLRQ 1129

Qy    810 YASEI---WPHMEKL---ESEVFSGEDVHPLFLQAISSALQF 845
            :   |  : |:      : |    :  ||:    |  |
Db   1130 RLQALDTGWNELHKMWENRQNLLSQSHAYQQFLRDTKQAEAF 1171


RESULT   2
US-08-968-751-2
; Sequence 2, Application US/08968751
; Patent No. 5948643
;  GENERAL INFORMATION:
;    APPLICANT:  Rubinfeld, Bonnee
;    APPLICANT:  Polakis, Paul G.
;    APPLICANT:  Ligenfelter, Carol
;    APPLICANT:  Vuong, Terilyn T.
;    TITLE OF INVENTION:  MODULATORS OF BRCA1 ACTIVITY
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ONYX Pharmaceuticals, Inc.
;      STREET:  3031 Research Drive
;      CITY:  Richmond
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94806
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/968,751
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Giotta, Gregory
;      REGISTRATION NUMBER:  32,028
;      REFERENCE/DOCKET NUMBER:  ONYX1024 GG
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 262-8710
;      TELEFAX:  (510) 222-9758
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 469 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-968-751-2


  Query Match             2.9%;  Score 125;  DB 2;  Length 469;
  Best Local Similarity   21.6%;  Pred. No. 0.0018;
  Matches   90;  Conservative   60;  Mismatches  180;  Indels   86;  Gaps   14;

Qy    392 VRALFEKISDLDMESRTRLILWFSH--HLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQ-- 447
          :|||    ||    ||    :   |  ||   | :  | |          || |   |
Db      3 IRALCTICSDFFDHSRDVAAIHCGHTFHL---QCLIQWFE---------TAPSRTCPQCR 50

Qy    448 -EILQREVRLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSR 506
           :: :| :    :  : |  ||    | |                  | |   : |:||:
Db     51 IQVGKRTIINKLFFDLAQEEENVLDAEFL----------------KNELDNVRAQLSQ 92

Qy    507 KVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSK 566
          | |||: ::                        ::: || |   :    :|:: :  |:
Db     93 KDKEKRDSQ--------------------VIIDTLRDTLEERNATVVSLQQALGKA-EM 130

Qy    567 LCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRM-----MGYRLVSNQAIVRWVFSPENVD 621
          ||    ||:  | |      |   :        :|      :   | | : | :     |
Db    131 LCSTLKKQMKYLEQQQDETKQAQEEARRLRSKMKTMEQIELLLQSQRPEVEEMIRDMGVG 190

Qy    622 QFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSL 681
          |  |  :   :  :| | |  : : ||        ||:|     | :| :: |||
Db    191 QSAV--EQLAVYCVSLKKEYENLKEARKASGE------VADK----LRKDLFSSRSKLQT 238

Qy    682 VEGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLFQSF 741
          |  |  | :   ::|  :  ::   :  :||:: |   :  ||     :| :  |: :|
Db    239 VYSE--LDQAKLELKSAQKDLQSADKEIMSLKKKLTMLQETLNLPPVASETVDRLVLES- 295

Qy    742 LGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCL 797
                | | :|       |   : |  :  |||:    |  | : |  |: ||
Db    296 --------PAPVEVNLKLRRPSFRDDIDLNATFDVDTPPARPSSS-QHGYYEKLCL 342


RESULT   3
US-08-795-475-6
; Sequence 6, Application US/08795475
; Patent No. 5965390
;  GENERAL INFORMATION:
;    APPLICANT:  Bjvrck, Lars
;    APPLICANT:  Sjvbring, Ulf
;    TITLE OF INVENTION:  PROTEIN L AND HYBRID PROTEINS THEREOF
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/795,475
;      FILING DATE:  11-FEB-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
```

```
;      REFERENCE/DOCKET NUMBER:  100084.402D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  443 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-795-475-6


  Query Match             2.8%;  Score 122;  DB 2;  Length 443;
  Best Local Similarity   20.0%;  Pred. No. 0.0032;
  Matches   90;  Conservative   83;  Mismatches  167;  Indels  110;  Gaps   20;

Qy    452 REVRLSYWDK-IKQSIENATALEELLPPKAGPNFMYS--LEEGKEKTEEQQLSAELSRKV 508
          : :|| : :| :|  :||| :       || :|  :  ||: |:  |:|:   :|  |:
Db     21 QNIRLRHENKDLKARLENAMEV-------AGRDFKRAEELEKAKQALEDQR--KDLETKL 71

Qy    509 KEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTL---LDIGSKSFTHLVTVLERYGQVFS 565
          || |   |:    | | :     |  |    : |   :|  |: : |  | ||
Db     72 KELQQDYDLAK--ESTSWDRQRLEKELEEKKEALELAIDQASRDY-HRATALE------- 121

Qy    566 KLCPDNDKQVML-LSQVS-TYWKNNV---QMTAVAIDRMMGYRLVSNQAIVRWVFSPEN- 619
          |   :  | : | : | |  | : ||   ::  :  :: :   |: |  :      | |
Db    122 KELEEKKKALELAIDQASQDYNRANVLEKELETITREQEINRNLLGNAKLELDQLSSEKE 181

Qy    620 ---------VDQFHVSDQPWEILGNALNKTYNRISDLRKDISNIT-------KNVLVAEK 663
                  ::  :||   : |   |: :       : ||::|:|       ::  :::
Db    182 QLTIEKAKLEEEKQISDASRQSLRRDLDASREAKKQVEKDLANLTAELDKVKEDKQISDA 241

Qy    664 ASANARVELEAAE-------------SKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAE 709
          :   | :|:|:               ::|  |: |  : :  |  :||:  :: : ||:
Db    242 SRQRLRRDLDASREAKKQVEKDLANLTAELDKVKEEKQISD--ASRQRLRRDLDASREAK 299

Qy    710 LSLRESLE------AKEALLNRALSETEVL-------LLLLFQSFLGVLKERLPDPTKVR 756
          : ::||       |    ||: | |:: |       |    ::    |||:|      :
Db    300 KQVEKALEEANSKLAALEKLNKELEESKKLTEKEKAELQAKLEAEAKALKEQL-----AK 354

Qy    757 SVQDLKSIGA--------EDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLTAFTR 808
            ::|  : |         | ||    |  :   |:   :  : :
Db    355 QAEELAKLRAGKASDSQTPDTKPGNKAVPGKGQAPQAGTKPNQNKA-------------- 400

Qy    809 QYASEIWPHMEKLESEVFSGEDVHPLFLQA 838
                |  |       :||  :| |  |
Db    401 -------PMKETKRQLPSTGETANPFFTAA 423


RESULT   4
US-09-052-089A-1
; Sequence 1, Application US/09052089A
; Patent No. 6346605
;   GENERAL INFORMATION:
;        APPLICANT: Lee, Soo Y.
;                   Choi, Yongwon
;        TITLE OF INVENTION: SIGNAL TRANSDUCER FOR THE TNF RECEPTOR SUPER
;                            FAMILY, AND USES THEREOF
;        NUMBER OF SEQUENCES: 16
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: David A. Jackson, Esq.
;             STREET: 411 Hackensack Ave, Continental Plaza, 4th
;                     Floor
;             CITY: Hackensack
;             STATE: New Jersey
;             COUNTRY: USA
;             ZIP: 07601
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/052,089A
;             FILING DATE: 31-Mar-1998
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Jackson Esq., David A.
;             REGISTRATION NUMBER: 26,742
;             REFERENCE/DOCKET NUMBER: 600-1-198 CIP 1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 201-487-5800
;             TELEFAX: 201-343-1684
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 469 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: NO
;        FRAGMENT TYPE: <Unknown>
;        ORIGINAL SOURCE:
;             ORGANISM: Homo sapiens
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-052-089A-1


  Query Match             2.7%;  Score 120;  DB 4;  Length 469;
  Best Local Similarity   20.7%;  Pred. No. 0.0055;
  Matches   87;  Conservative   65;  Mismatches  174;  Indels   94;  Gaps   14;

Qy    392 VRALFEKISDLDMESRTRLILWFSH--HLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQ-- 447
          :|||    ||    ||     :   |  ||                ::   : || |   |
Db      3 IRALCTICSDFFDHSRDVAAIHCGHTFHLQ-----------CLIQSFETAPSRTCPQCR 50

Qy    448 -EILQREVRLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSR 506
           :: :| :    :  : |  ||    | |                  | |   : |:||:
Db     51 IQVGKRTIINKLFFDLAQEEENVLDREFL-----------------KNELDNVRAQLSQ 92

Qy    507 KVKEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSK 566
          | |||: ::                       ::: || |   :    :|:: :  |:
Db     93 KDKEKRDSQ--------------------VIIDTLRDTLEERNATVVSLQQALGKA-EM 130

Qy    567 LCPDNDKQVMLLSQVSTYWK---------NNVQMTAVAIDRMMGYRLVSNQAIVRWVFSP 617
          ||    ||: | |      |           :   |   |: ::  :|   : ::|
Db    131 LCSTLKKQMKYLEQQQDETKQAQEEAGRLRSKMKTMEQIELLLQSQLPEVEEMIR----D 186

Qy    618 ENVDQFHVSDQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAES 677
            | |  |  :   :  :| | |  : : ||          ||:|     | :| :: |
Db    187 MGVGQSAV--EQLAVYCVSLKKEYENLKEARKASGE------VADK----LRKDLFSSRS 234

Qy    678 KLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLL 737
          ||  |  |  | :   ::| ::  ::    :  :||:: |   :  ||     :| :  |:
Db    235 KLQTVYSE--LDQAKLELKSAQKDLQSADKEIMSLKKKLTMLQETLNLPPVASETVDRLV 292

Qy    738 FQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCL 797
            :|         | | :|       | ' : |  :  |||:'   |  | : |  |: ||
Db    293 LES---------PAPVEVNLKLRRPSFRDDIDLNATFDVDTPPARPSSS-QHGYYEKLCL 342


RESULT   5
US-09-019-160-2
; Sequence 2, Application US/09019160
; Patent No. 6306588
;  GENERAL INFORMATION:
;    APPLICANT:  Chatterjee, Deb K.
;    APPLICANT:  Solus, Joseph
;    APPLICANT:  Yang, Shuwei
;    TITLE OF INVENTION:  Polymerases for Analyzing or Typing Polymorphic
;    TITLE OF INVENTION:  Nucleic Acid Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C
;      STREET:  1100 New York Ave., N.W., Suite 600
```

```
;       CITY:  Washington
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20005-3934
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/019,160
;       FILING DATE:  06-FEB-1998
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  (To be assigned)
;       FILING DATE:  06-JAN-1998
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/037,393
;       FILING DATE:  07-FEB-1997
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Esmond, Robert W.
;       REGISTRATION NUMBER:  32,893
;       REFERENCE/DOCKET NUMBER:  0942.4250002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-371-2600
;       TELEFAX:  202-371-2540
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  893 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  protein
US-09-019-160-2


  Query Match             2.7%;  Score 120;  DB 4;  Length 893;
  Best Local Similarity   18.6%;  Pred. No. 0.017;
  Matches   99;  Conservative   92;  Mismatches  188;  Indels  154;  Gaps   22;

Qy    422 QFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSY---------------WDKIK--- 463
          | :  :    ::|:  :  |:||    :  ||| :                  |:::|
Db    206 QLLGKYRNLEYILEHARELPQRVRKALLRDREVAILSKKLATLVTNAPVEVDWEEMKYRG 265

Qy    464 ----------QSIENATALEEL-LPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQ 512
                   : :| |: ::|| |  :| |        | |  :: :   :|  |:||
Db    266 YDKRKLLPILKELEFASIMKELQLYEEAEPT-------GYEIVKDHKTFEDLIEKLKEVP 318

Qy    513 T-ARDM-----------IVWIEETIYPVHGFEVTL----------TIVVQTLLDIGSKSF 550
          : | |:           || |  :  |   : : |          |:|:  | :|
Db    319 SFALDLETSSLDPFNCEIVGISVSFKPKTAYYIPLHHRNAQNLDETLVLSKLKEILEDPS 378

Qy    551 THLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKN-NVQMTAVAID------------ 597
          : :|    :|           | :|:::  :|  : : :  : |  ::
Db    379 SKIVGQNLKY----------DYKVLMVKGISPVYPHFDTMIAAYLLEPNEKKFNLEDLS 427

Qy    598 -RMMGYRLVSNQAIVRWV-----------------FSPENVDQFHVSDQPWEILGNALN 638
           : ::||:: | | :: :                  :| |: |   :: : ::||   |:
Db    428 LKFLGYKMTSYQELMSFSSPLFGFSFADVPVDKAANYSCEDAD---ITYRLYKILSMKLH 484

Qy    639 K-----TYNRISD------LRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPV 687
          :      : ||        | ::: :  :    :| |     :||    |:  : |||
Db    485 EAELENVFYRIEMPLVNVLARMELNGVYVDTEFLKKLSEEYGKKLEELAEKIYQIAGEPF 544

Qy    688 LGENPAKMKRL---------KSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLF 738
             :| :: ::          :    |||  | | |  :|  | : |    | |:: |:|
Db    545 NINSPKQVSKILFEKLGIKPRGKTTKTG--EYSTR--IEVLEEIAN----EHEIVPLILE 596

Qy    739 QSFLGVLKE-------RLPDPTKVRSVQDLKSIGAEDDKPSAMD-------VDSENGNPK 784
          :  ||         :| :|   |       |    : |: |           || |
Db    597 YRKIQKLKSTYIDTLPKLVNPKTGRIHASFHQTGTATGRLSSSDPNLQNLPTKSEEGKEI 656

Qy    785 KSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHM---EKLESEVFSGEDVHPL 834
          :   | :   | : :  |      |    |  |:   | |      | ||| |
Db    657 RKAIVPQDPDWWIVSADY-----SQIELRILAHLSGDENLVKAFEEGIDVHTL 704


RESULT   6
US-07-853-913-2
; Sequence 2, Application US/07853913
; Patent No. 5338839
;  GENERAL INFORMATION:
;    APPLICANT:  McKay, Ronald D.G.
;    APPLICANT:  Lendahl, Urban
;    TITLE OF INVENTION:  Nestin Expression As An Indicator of
;    TITLE OF INVENTION:  Neuroepithelial Tumors
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/853,913
;      FILING DATE:  19920319
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/660,412
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/603,803
;      FILING DATE:  25-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/201,762
;      FILING DATE:  02-JUN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/180,548
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-4641AAAA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1805 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-853-913-2


  Query Match             2.7%;  Score 119;  DB 1;  Length 1805;
  Best Local Similarity   20.8%;  Pred. No. 0.077;
  Matches  150;  Conservative  110;  Mismatches  241;  Indels  220;  Gaps   37;

Qy    218 EDLLDRIQSLASNGW--KLES-VPRPHLSFEAQLVAGKFHELRPIK--------CMEQPS 266
          :| |  :: |    |  |||: | | :|: | : |||:  ::|  :        |  :
Db     57 DDELASLRILVDQRWREKLEAEVQRDNLAEELESVAGRCQQVRLARERTVQEAACSRRAL 116

Qy    267 PPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDVLFYLNG 326
              ::| :   |  : | |  : :|      |:: |                    :||
Db    117 EAEKNARGWLSTQAAE-LERELEALR-----AAHEEERA----------------HLN- 152

Qy    327 CRKECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPA 386
          :  ||                          | || :              :||  |
```

```
Db    153 AQAACAPRRP----------------------PAPPHR--------------IPGPAPE 175

Qy    387 VV----------AGAVRALFEKISDLDM---ESRTRLILWFSHHLSNFQFIWPWEEWAFV 433
Db    176 VEDLARRLGEVWRGAVRDYQERVAHMESSLGQARERL----------SQAVRGARECR-- 223

Qy    434 LDLPKWAPKRVFVQ---EILQREVRLSYWDKIKQSIENATALEELLPPKAG--PNFMYSL 488
Db    224 LEVQQLQADRDSLQERREALEQRLEGRWQDRLQATDKFQLAVEALEQEKQGLQSQIAQIL 283

Qy    489 EEGKE-----------KTEEQQLSAELSR----------------KVK-------EKQ 512
Db    284 EGGQQLAHLKMSLSLEVATYRTLLEAENSRLQTPGRGSQASLGFLDPKLKPNFLGIPEDQ 343

Qy    513 TARDMIVWIEETIYP---VHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCP 569
Db    344 YLGSVLPALSPTSFPSPLPNTLETPVTAFLKTQEFLQARTPTLASTPIP---PISEAPCP 400

Qy    570 DNDK---QVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVS 626
Db    401 PNAEVRAQEVPLSLLQTQAPEPLWLKAT---------VPSSSAILPELEEPGGKQQGHFP 451

Qy    627 DQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLV---- 682
Db    452 D----------------------DLTSLATN-LNPHHPTLEAK-DGESSESRVSSIFQED 487

Qy    683 EG---EPVLGENPAKMKRLKSTVEKTGEAELSLRESLEA-----KEALLNRALSETEVLL 734
Db    488 EGQIWELVEKEADIEVKVENSSAQKTQESGLDTEETQDSQGPLQKETL--KALGE-EPLM 544

Qy    735 LLLFQSFLGVLKER-----------LPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNP 783
Db    545 SLKIQNYETAGKENCNSSTEGHLGTLEGPEKEKQI-PLKSL---EEKNVESEKTLENGVP 600

Qy    784 KKSCEVGE-----REQWCLSTLGYLTAFT---RQYASEIWPHME-KLES-EVFSGEDVHP 833
Db    601 VLSELLGKEDTRTEDQELMSPKGTLKRFSSLGKESQEVVRPSKEGNLESWTAFKEESQHP 660

Qy    834 L 834
Db    661 L 661


RESULT   7
US-09-019-160-9
; Sequence 9, Application US/09019160
; Patent No. 6306588
;  GENERAL INFORMATION:
;    APPLICANT:  Chatterjee, Deb K.
;    APPLICANT:  Solus, Joseph
;    APPLICANT:  Yang, Shuwei
;    TITLE OF INVENTION:  Polymerases for Analyzing or Typing Polymorphic
;    TITLE OF INVENTION:  Nucleic Acid Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C
;      STREET:  1100 New York Ave., N.W., Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/019,160
;      FILING DATE:  06-FEB-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  (To be assigned)
;      FILING DATE:  06-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/037,393
;      FILING DATE:  07-FEB-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0942.4250002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  893 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
US-09-019-160-9


  Query Match             2.7%;  Score 117.5;  DB 4;  Length 893;
  Best Local Similarity   18.4%;  Pred. No. 0.031;
  Matches  132;  Conservative  117;  Mismatches  244;  Indels  225;  Gaps   31;

Qy    298 PANKMEDV-QPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFRYEYLM---AETLF 353
Db     31 PTNAVYDVARMLVKFIKEHIIPEKDYAAVAFDKKAA--------TFRHKLLVSDKAQRPK 82

Qy    354 SQILLLPQPPF-KTLYYTLVIMDLCKALPG--AFPAVVAGAVRA---------------- 394
Db     83 TPALLVQQLPYIKRLIEALGFKVL--ELEGYEADDIIATLAVRAARFLMRFSLITGDKDM 140

Qy    395 ------------LFEKISDLDMESRTR--------------LILWFSHHLSNF------- 421
Db    141 LQLVNEKIKVWRIVKGISDLELYDSKKVKERYGVEPHQIPDLLALTGDDIDNIPGVTGIG 200

Qy    422 -----QFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSY---------------WDK 461
Db    201 EKTAVQLLGKYRNLEYILEHARELPQRVRKALLRDREVAILSKKLATLVTNAPVEVDWEE 260

Qy    462 IK------------QSIENATALEEL-LPPKAGPNFMYSLEEGKEKTEEQQLSAELSRK 507
Db    261 MKYRGYDKRKLLPILKELEFASIMKELQLYEEAEPT-------GYEIVKDHKTFEDLIEK 313

Qy    508 VKE--------KQTARD----MIVWIEETIYPVHGFEVTL----------TIVVQTLLDI 545
Db    314 LKEVPSFALALETSSLDPFNCEIVGISVSFKPKTAYYIPLHHRNAQNLDETLVLSKLKEI 373

Qy    546 GSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKN-NVQMTAVAID------- 597
Db    374 LEDPSSKIVGQNLKY----------DYKVLMVKGISPVYPHFDTMIAAYLLEPNEKKFN 422

Qy    598 ------RMMGYRLVSNQAIVRWV----------------FSPENVDQFHVSDQPWEIL 633
Db    423 LEDLSLKFLGYKMTSYQELMSFSSPLFGFSFADVPVDKAANYSCEDAD---ITYRLYKIL 479

Qy    634 GNALNK-----TYNRISD------LRKDISNITKNVLVAEKASANARVELEAAESKLSLV 682
Db    480 SMKLHEAELENVFYRIEMPLVNVLARMELNGVYVDTEFLKKLSEEYGKKLEELAEKIYQI 539

Qy    683 EGEPVLGENPAKMKRL---------KSTVEKTGEAELSLRESLEAKEALLNRALSETEVL 733
Db    540 AGEPFNINSPKQVSKILFEKLGIKPRGKTTKTG--EYSTR--IEVLEEIAN----EHEIV 591

Qy    734 LLLLFQSFLGVLKE-------RLPDPTKVRSVQDLKSIGAEDDKPSAMD-------VDSE 779
Db    592 PLILEYRKIQKLKSTYIDTLPKLVNPKTGRIHASFHQTGTATGRLSSSDPNLQNLPTKSE 651

Qy    780 NGNPKKSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHM---EKLESEVFSGEDVHPL 834
Db    652 EGKEIRKAIVPQDPDWWIVSADY-----SQIELRILAHLSGDENLVKAFEEGIDVHTL 704
```

```
RESULT   8
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT: DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 0660-0125-0 PCT
;  CURRENT APPLICATION NUMBER: US/08/973,462B
;  CURRENT FILING DATE: 1998-02-06
;  EARLIER APPLICATION NUMBER: PCT/FR96/00894
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: FR 95/07007
;  EARLIER FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 1786
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Polypeptide
US-08-973-462-8


  Query Match             2.7%;  Score 116.5;  DB 4;  Length 1786;
  Best Local Similarity   18.2%;  Pred. No. 0.13;
  Matches   84;  Conservative   83;  Mismatches  152;  Indels  143;  Gaps   18;

Qy    446 VQEILQREVRLSYWDKIKQSI-EN------------ATALEELLPPKAGPNFMYSLEEG 491
          |:| :   |  |  : :::|: ||             |  :||:: |    :   |:||
Db    699 VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVEES 758

Qy    492 KEKTEEQQLSAELSRKVKE-------KQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLD 544
          |:  |: ::  :   | |       :    :   :|| : |     |  ::      :
Db    759 VEENVEESVAENVEESVAENVEESVAENVEESVAPTVEEIVAPSVEESVAPSVEESVAEN 818

Qy    545 IGSKSFTHLVTVLERYGQVFSKLCPD---NDKQVMLLSQVSTYWKNNVQMTAVAIDRMMG 601
          : :    :|:: |   | : ::   |   |: : :  : |:|  :|  : || ::
Db    819 VATNLSDNLLSNL--LGGIETEEIKDSILNEIEEVKENVVTTILENVEETTAESVTTFSN 876

Qy    602 Y------RLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNR---------ISD 646
                ::|  |       | ::: | :     :|  ||  | :             | :
Db    877 ILEEIQENTITNDTI------EEKLEELHEN-----VLSAALENTQSEEEKKEVIDVIEE 925

Qy    647 LRKDI-SNITKNVLVAEKASANARVEL------EAAESKLSLVEGEPVLGEN--PAKMKR 697
          ::::: : : : |  ||: |||   |:       | ||  :: |    | |      : :
Db    926 VKEEVATTLIETVEQAEEKSANTITEIFENLEENAVESNENVAENLEKLNETVFNTVLDK 985

Qy    698 LKSTVEKTGEA----------------------------------------ELSLRE 714
          :: ||| :||:                                        :: | |
Db    986 VEETVEISGESLENNEMDKAFFSEIFDNVKGIQENLLTGMFRSIETSIVIQSEEKVDLNE 1045

Qy    715 SL---------EAKEALLNRA--LSETE----------------VLLLLLFQSFLGVLK 746
          ::           || |||:   :| ||                | :  :   |||:|
Db   1046 NVVSSILDNIENMKEGLLNKLENISSTEGVQETVTEHVEQNVYVDVDVPAMKDQFLGILN 1105

Qy    747 E---------RLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSE 779
          |           | |  |  |  |: ::    |:|   :|: |
Db   1106 EAGGLKEMFFNLEDVFK--SESDVITVEEIKDEPVQKEVEKE 1145


RESULT   9
US-09-019-160-5
; Sequence 5, Application US/09019160
; Patent No. 6306588
;  GENERAL INFORMATION:
;    APPLICANT:  Chatterjee, Deb K.
;    APPLICANT:  Solus, Joseph
;    APPLICANT:  Yang, Shuwei
;    TITLE OF INVENTION:  Polymerases for Analyzing or Typing Polymorphic
;    TITLE OF INVENTION:  Nucleic Acid Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C
;      STREET:  1100 New York Ave., N.W., Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/019,160
;      FILING DATE:  06-FEB-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  (To be assigned)
;      FILING DATE:  06-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/037,393
;      FILING DATE:  07-FEB-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0942.4250002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  708 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
US-09-019-160-5


  Query Match             2.7%;  Score 116;  DB 4;  Length 708;
  Best Local Similarity   18.4%;  Pred. No. 0.029;
  Matches   98;  Conservative   93;  Mismatches  188;  Indels  154;  Gaps   22;

Qy    422 QFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSY--------------WDKIK--- 463
          | :  :    ::|:  :  |:||    :  ||| :                 |:::|
Db     21 QLLGKYRNLEYILEHARELPQRVRKALLRDREVAILSKKLATLVTNAPVEVDWEEMKYRG 80

Qy    464 ----------QSIENATALEEL-LPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKE-- 510
                   : :| |: ::|| |  :| |         | |  :: :   :|  |:||
Db     81 YDKRKLLPILKELEFASIMKELQLYEEAEPT-------GYEIVKDHKTFEDLIEKLKEVP 133

Qy    511 ------KQTARD----MIVWIEETIYPVHGFEVTL----------TIVVQTLLDIGSKSF 550
                : :: |      || |  :  |   : : |          |:|:  | :|
Db    134 SFALALETSSLDPFNCEIVGISVSFKPKTAYYIPLHHRNAQNLDETLVLSKLKEILEDPS 193

Qy    551 THLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKN-NVQMTAVAID------------ 597
          : :|    :|           | :|:::  :|  : : :  : |  ::
Db    194 SKIVGQNLKY----------DYKVLMVKGISPVYPHFDTMIAAYLLEPNEKKFNLEDLS 242

Qy    598 -RMMGYRLVSNQAIVRWV----------------FSPENVDQFHVSDQPWEILGNALN 638
           : :||:: | | :: :                 :| |: |   :: : ::||    |:
Db    243 LKFLGYKMTSYQELMSFSSPLFGFSFADVPVDKAANYSCEDAD---ITYRLYKILSMKLH 299

Qy    639 K-----TYNRISD------LRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPV 687
          :      : ||         | ::: :  :    :| |     :||    |:  : |||
Db    300 EAELENVFYRIEMPLVNVLARMELNGVYVDTEFLKKLSEEYGKKLEELAEKIYQIAGEPF 359

Qy    688 LGENPAKMKRL---------KSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLF 738
```

```
Db    360 NINSPKQVSKILFEKLGIKPRGKTTKTG--EYSTR--IEVLEEIAN----EHEIVPLILE 411

Qy    739 QSFLGVLKE-------RLPDPTKVRSVQDLKSIGAEDDKPSAMD-------VDSENGNPK 784

Db    412 YRKIQKLKSTYIDTLPKLVNPKTGRIHASFHQTGTATGRLSSSDPNLQNLPTKSEEGKEI 471

Qy    785 KSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHM---EKLESEVFSGEDVHPL 834

Db    472 RKAIVPQDPDWWIVSADY-----SQIELRILAHLSGDENLVKAFEEGIDVHTL 519


RESULT 10
US-09-019-160-6
; Sequence 6, Application US/09019160
; Patent No. 6306588
;  GENERAL INFORMATION:
;    APPLICANT:  Chatterjee, Deb K.
;    APPLICANT:  Solus, Joseph
;    APPLICANT:  Yang, Shuwei
;    TITLE OF INVENTION:  Polymerases for Analyzing or Typing Polymorphic
;    TITLE OF INVENTION:  Nucleic Acid Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C
;      STREET:  1100 New York Ave., N.W., Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/019,160
;      FILING DATE:  06-FEB-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  (To be assigned)
;      FILING DATE:  06-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/037,393
;      FILING DATE:  07-FEB-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0942.4250002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  893 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-019-160-6


  Query Match             2.7%;  Score 116;  DB 4;  Length 893;
  Best Local Similarity   18.4%;  Pred. No. 0.043;
  Matches   98;  Conservative   93;  Mismatches  188;  Indels  154;  Gaps   22;

Qy    422 QFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSY--------------WDKIK--- 463

Db    206 QLLGKYRNLEYILEHARELPQRVRKALLRDREVAILSKKLATLVTNAPVEVDWEEMKYRG 265

Qy    464 ----------QSIENATALEEL-LPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKE-- 510

Db    266 YDKRKLLPILKELEFASIMKELQLYEEAEPT-------GYEIVKDHKTFEDLIEKLKEVP 318

Qy    511 ------KQTARD----MIVWIEETIYPVHGFEVTL----------TIVVQTLLDIGSKSF 550

Db    319 SFALALETSSLDPFNCEIVGISVSFKPKTAYYIPLHHRNAQNLDETLVLSKLKEILEDPS 378

Qy    551 THLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKN-NVQMTAVAID------------ 597

Db    379 SKIVGQNLKY----------DYKVLMVKGISPVYPHFDTMIAAYLLEPNEKKFNLEDLS 427

Qy    598 -RMMGYRLVSNQAIVRWV-----------------FSPENVDQFHVSDQPWEILGNALN 638

Db    428 LKFLGYKMTSYQELMSFSSPLFGFSFADVPVDKAANYSCEDAD---ITYRLYKILSMKLH 484

Qy    639 K-----TYNRISD------LRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPV 687

Db    485 EAELENVFYRIEMPLVNVLARMELNGVYVDTEFLKKLSEEYGKKLEELAEKIYQIAGEPF 544

Qy    688 LGENPAKMKRL---------KSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLF 738

Db    545 NINSPKQVSKILFEKLGIKPRGKTTKTG--EYSTR--IEVLEEIAN----EHEIVPLILE 596

Qy    739 QSFLGVLKE-------RLPDPTKVRSVQDLKSIGAEDDKPSAMD-------VDSENGNPK 784

Db    597 YRKIQKLKSTYIDTLPKLVNPKTGRIHASFHQTGTATGRLSSSDPNLQNLPTKSEEGKEI 656

Qy    785 KSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHM---EKLESEVFSGEDVHPL 834

Db    657 RKAIVPQDPDWWIVSADY-----SQIELRILAHLSGDENLVKAFEEGIDVHTL 704


RESULT 11
US-09-019-160-7
; Sequence 7, Application US/09019160
; Patent No. 6306588
;  GENERAL INFORMATION:
;    APPLICANT:  Chatterjee, Deb K.
;    APPLICANT:  Solus, Joseph
;    APPLICANT:  Yang, Shuwei
;    TITLE OF INVENTION:  Polymerases for Analyzing or Typing Polymorphic
;    TITLE OF INVENTION:  Nucleic Acid Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C
;      STREET:  1100 New York Ave., N.W., Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/019,160
;      FILING DATE:  06-FEB-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  (To be assigned)
;      FILING DATE:  06-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/037,393
;      FILING DATE:  07-FEB-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0942.4250002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
```

```
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  893 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
US-09-019-160-7


  Query Match             2.7%;  Score 116;  DB 4;  Length 893;
  Best Local Similarity   18.4%;  Pred. No. 0.043;
  Matches   98;  Conservative   93;  Mismatches  188;  Indels  154;  Gaps   22;

Qy     422 QFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSY--------------WDKIK--- 463
           | :  :    ::|:  :  |:||    :  ||| :                  |:::|
Db     206 QLLGKYRNLEYILEHARELPQRVRKALLRDREVAILSKKLATLVTNAPVEVDWEEMKYRG 265

Qy     464 ----------QSIENATALEEL-LPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKE-- 510
                    : :| |: ::|| |  :| |        | |  :: :   :|  |:||
Db     266 YDKRKLLPILKELEFASIMKELQLYEEAEPT-------GYEIVKDHKTFEDLIEKLKEVP 318

Qy     511 ------KQTARD----MIVWIEETIYPVHGFEVTL----------TIVVQTLLDIGSKSF 550
                : :: |     || |  :  |   : : |          |:|:  | :|
Db     319 SFALALETSSLDPFNCEIVGISVSFKPKTAYYIPLHHRNAQNLDETLVLSKLKEILEDPS 378

Qy     551 THLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKN-NVQMTAVAID------------ 597
           : :|    :|           | :|:::  :|  : : :  : |  ::
Db     379 SKIVGQNLKY----------DYKVLMVKGISPVYPHFDTMIAAYLLEPNEKKFNLEDLS 427

Qy     598 -RMMGYRLVSNQAIVRWV-----------------FSPENVDQFHVSDQPWEILGNALN 638
            : :||:: | | :: :                  :| |: |   :: : ::||   |:
Db     428 LKFLGYKMTSYQELMSFSSPLFGFSFADVPVDKAANYSCEDAD---ITYRLYKILSMKLH 484

Qy     639 K-----TYNRISD------LRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPV 687
           :      : ||         | ::: :  :    :| |     :||    |:  : |||
Db     485 EAELENVFYRIEMPLVNVLARMELNGVYVDTEFLKKLSEEYGKKLEELAEKIYQIAGEPF 544

Qy     688 LGENPAKMKRL---------KSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLF 738
              :| :: ::         :    |||  | | |  :|  | : |    | |:: |:|
Db     545 NINSPKQVSKILFEKLGIKPRGKTTKTG--EYSTR--IEVLEEIAN----EHEIVPLILE 596

Qy     739 QSFLGVLKE-------RLPDPTKVRSVQDLKSIGAEDDKPSAMD-------VDSENGNPK 784
           :  ||        :| :|   |         |    : |: |          || |
Db     597 YRKIQKLKSTYIDTLPKLVNPKTGRIHASFHQTGTATGRLSSSDPNLQNLPTKSEEGKEI 656

Qy     785 KSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHM---EKLESEVFSGEDVHPL 834
           :  | :   | : :  |       |   |  |:   | |       | |||  |
Db     657 RKAIVPQDPDWWIVSADY-----SQIELRILAHLSGDENLVKAFEEGIDVHTL 704


RESULT  12
US-09-019-160-8
; Sequence 8, Application US/09019160
; Patent No. 6306588
;  GENERAL INFORMATION:
;    APPLICANT:  Chatterjee, Deb K.
;    APPLICANT:  Solus, Joseph
;    APPLICANT:  Yang, Shuwei
;    TITLE OF INVENTION:  Polymerases for Analyzing or Typing Polymorphic
;    TITLE OF INVENTION:  Nucleic Acid Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C
;      STREET:  1100 New York Ave., N.W., Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/019,160
;      FILING DATE:  06-FEB-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  (To be assigned)
;      FILING DATE:  06-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/037,393
;      FILING DATE:  07-FEB-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0942.4250002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  893 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
US-09-019-160-8


  Query Match             2.7%;  Score 116;  DB 4;  Length 893;
  Best Local Similarity   18.4%;  Pred. No. 0.043;
  Matches   98;  Conservative   93;  Mismatches  188;  Indels  154;  Gaps   22;

Qy     422 QFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSY--------------WDKIK--- 463
           | :  :    ::|:  :  |:||    :  ||| :                  |:::| .
Db     206 QLLGKYRNLEYILEHARELPQRVRKALLRDREVAILSKKLATLVTNAPVEVDWEEMKYRG 265

Qy     464 ----------QSIENATALEEL-LPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKE-- 510
                    : :| |: ::|| |  :| |        | |  :: :   :|  |:||
Db     266 YDKRKLLPILKELEFASIMKELQLYEEAEPT-------GYEIVKDHKTFEDLIEKLKEVP 318

Qy     511 ------KQTARD----MIVWIEETIYPVHGFEVTL----------TIVVQTLLDIGSKSF 550
                : :: |     || |  :  |   : : |          |:|:  | :|
Db     319 SFALALETSSLDPFNCEIVGISVSFKPKTAYYIPLHHRNAQNLDETLVLSKLKEILEDPS 378

Qy     551 THLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKN-NVQMTAVAID------------ 597
           : :|    :|           | :|:::  :|  : : :  : |  ::
Db     379 SKIVGQNLKY----------DYKVLMVKGISPVYPHFDTMIAAYLLEPNEKKFNLEDLS 427

Qy     598 -RMMGYRLVSNQAIVRWV-----------------FSPENVDQFHVSDQPWEILGNALN 638
            : :||:: | | :: :                  :| |: |   :: : ::||   |:
Db     428 LKFLGYKMTSYQELMSFSSPLFGFSFADVPVDKAANYSCEDAD---ITYRLYKILSMKLH 484

Qy     639 K-----TYNRISD------LRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPV 687
           :      : ||         | ::: :  :    :| |     :||    |:  : |||
Db     485 EAELENVFYRIEMPLVNVLARMELNGVYVDTEFLKKLSEEYGKKLEELAEKIYQIAGEPF 544

Qy     688 LGENPAKMKRL---------KSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLLLF 738
              :| :: ::         :    |||  | | |  :|  | : |    | |:: |:|
Db     545 NINSPKQVSKILFEKLGIKPRGKTTKTG--EYSTR--IEVLEEIAN----EHEIVPLILE 596

Qy     739 QSFLGVLKE-------RLPDPTKVRSVQDLKSIGAEDDKPSAMD-------VDSENGNPK 784
           :  ||        :| :|   |         |    : |: |          || |
Db     597 YRKIQKLKSTYIDTLPKLVNPKTGRIHASFHQTGTATGRLSSSDPNLQNLPTKSEEGKEI 656

Qy     785 KSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHM---EKLESEVFSGEDVHPL 834
           :  | :   | : :  |       |   |  |:   | |       | |||  |
Db     657 RKAIVPQDPDWWIVSADY-----SQIELRILAHLSGDENLVKAFEEGIDVHTL 704


RESULT  13
```

```
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
;  GENERAL INFORMATION:
;    APPLICANT:  YEN, TIMOTHY J.
;    APPLICANT:  RATTNER, JEROME B.
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING A
;    TITLE OF INVENTION:  TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DANN, DORFMAN, HERRELL AND SKILLMAN
;      STREET:  1601 MARKET STREET, SUITE 720
;      CITY:  PHILADELPHIA
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/353,700
;      FILING DATE:  09-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  REED, JANET E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3248 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HUMAN
US-08-353-700-1


  Query Match             2.6%;  Score 113.5;  DB 1;  Length 3248;
  Best Local Similarity   18.7%;  Pred. No. 0.77;
  Matches  187;  Conservative  139;  Mismatches  332;  Indels  341;  Gaps   45;

Qy      9 LRIGEKGPEYGTSSD-YKDHIETCFGVIRREIERSGDQVLPFLLQC----------A 54
          | | ||  |    :|  || :|     : ||:: | :     :|
Db   2232 LHIAEKLKERERENDSLKDKVEN----LERELQMSEENQELVILDAENSKAEVETLKTQI 2287

Qy     55 EQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLC 114
          |::   : ::   :  | | |:  ::: | .          |  : :  ||    |||
Db   2288 EEMARSLKIFELDLVTLRSEKENLTKQIQEK----------QGQLSELDKLLSSFKSLLE 2337

Qy    115 SKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAE--- 171
           |  : |      |  :   : |  |   |   :        :|       |  |: :
Db   2338 EK--EQA------EIQIKEESKTAVEMLQNQLKELNEAVAALC-------GDQEIMKATE 2382

Qy    172 ----------QVPDEIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDL 220
                    |: : ||:    ::| |   :       |       |  : |  |  ||:|
Db   2383 QSLDPPIEEEHQLRNSIEK----LRARLEADEKKQLCVLQQLKESEHHADLL-KGRVENL 2437

Qy    221 LDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELR-PIKCMEQPSPPSDHSRAYSGKQ 279
            ::   :|        : | : ||:   |:   |:  |: | |      :         :
Db   2438 ERELEIARTN--------QEHAALEAENSKGEVETLKAKIEGMTQSLRGLELDVVTIRSE 2489

Qy    280 KH---DALTRYPQRIRRL--------NIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCR 328
          |    : | :  :||  |         ||    : | ||      ::|     :  |
Db   2490 KENLTNELQKEQERISELEIINSSFENILQEKEQEKVQ------MKEKSSTAMEMLQTQL 2543

Qy    329 KECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPA-- 386
          ||    :| |            : | ||:  |               ::  : | |   |
Db   2544 KELNERVAALHNDQEACKAKEQNLSSQVECLE------------LEKAQLLQGLDEAKN 2590

Qy    387 ---VVAGAVRALFEKISD----------------LDMESRTRLI-----------LWFSH 416
             |:  :|: | ::: |                  :: : :|:           ||
Db   2591 NYIVLQSSVKGLIQEVEDGKQKLEKKDEEISRLKNQIQDQEQLVSKLSQVEGEHQLWKEQ 2650

Qy    417 HLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEI----------LQREVRLSYWDKIK--Q 464
          :|          |:   ||       |   :|:           |: |: |:  ||:   :
Db   2651 NLELRNLTVELEQKIQVLQ-----SKNASLQDTLEVLQSSYKNLENELELTKMDKMSFVE 2705

Qy    465 SIENATALEELLPPKAGPNFMYSLEEGKEKT---------EEQQLSAELSRKVKEKQTAR 515
           :   || |  |           : |  :||          |: :|: ||   ::| ::::
Db   2706 KVNKMTAKETEL--------QREMHEMAQKTAELQEELSGEKNRLAGELQLLLEEIKSSK 2757

Qy    516 DMIVWIEETIYPVHGFEVTL-TIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLC------ 568
          | :              |:||    ::  ||   |         :|: |:|  ::
Db   2758 DQLK-----------ELTLENSELKKSLDCMHKD------QVEKEGKVREEIAEYQLRL 2799

Qy    569 --PDNDKQVMLLS-------QVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPEN 619
             :   | :||         :: || :                :| | :
Db   2800 HEAEKKHQALLLDTNKQYEVEIQTYRE---------------KLTSKEEC--------- 2834

Qy    620 VDQFHVSDQPWEI---------LGNALNKTYNRISDLRKDISNITK--NVLVAEKASANA 668
             :| |  ||          | |:|  |   : :|:|   :  |  | |  |
Db   2835 -----LSSQKLEIDLLKSSKEELNNSLKATTQILEELKKTKMDNLKYVNQLKKENERAQG 2889

Qy    669 RVEL------------EAAESKLSLVEG----------EPVLGENPAKMKRLKSTV-EK 704
          :::|             |  : :|| ::            :  : |   ::| || |: ||
Db   2890 KMKLLIKSCKQLEEEKEILQKELSQLQAAQEKQKTGTVMDTKVDELTTEIKELKETLEEK 2949

Qy    705 TGEAE---------LSLRESLE-AKEALLNRALSETEVLLLLLFQS--------FLGVLK 746
          | ||:          |   | || ||| |      ||:|  |   ||           || :
Db   2950 TKEADEYLDKYCSLLISHEKLEKAKEML------ETQVAHLCSQQSKQDSRGSPLLGPV- 3002

Qy    747 ERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLTAF 806
            :| |: : || : | : :  :| |
Db   3003 --VPGPSPIPSVTE-KRLSSGQNKASG------------------------------- 3026

Qy    807 TRQYASEIWPH------------MEKLESEVFSGEDVHP 833
           || :| || :              :| :  | ||  :||
Db   3027 KRQRSSGIWENGRGPTPATPESFSKKSKKAVMSG--IHP 3063


RESULT  14
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
;  GENERAL INFORMATION:
;    APPLICANT:  Yen, Timothy J.
;    APPLICANT:  Rattner, Jerome B.
;    TITLE OF INVENTION:  Nucleic Acid Encoding a Transiently
;    TITLE OF INVENTION:  Expressed Kinetochore Protein, and Methods of Use
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street   Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16216
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/353,700
```

```
;      FILING DATE:  09-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3248 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US95-16216-1
```

```
  Query Match             2.6%;  Score 113.5;  DB 5;  Length 3248;
  Best Local Similarity   18.7%;  Pred. No. 0.77;
  Matches  187;  Conservative  139;  Mismatches  332;  Indels  341;  Gaps   45;

Qy       9 LRIGEKGPEYGTSSD-YKDHIETCFGVIRREIERSGDQVLPFLLQC------------A 54
           | | ||  |    :|  || :|    : ||:: | :     :|
Db    2232 LHIAEKLKERERENDSLKDKVEN----LERELQMSEENQELVILDAENSKAEVETLKTQI 2287

Qy      55 EQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLC 114
           |::   : ::   :  |  | |:  ::: |           |  : :  ||    |||
Db    2288 EEMARSLKIFELDLVTLRSEKENLTKQIQEK----------QGQLSELDKLLSSFKSLLE 2337

Qy     115 SKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAE--- 171
           |  : |      |  :   : | |   |   :       :|       |  |: :
Db    2338 EK--EQA------EIQIKEESKTAVEMLQNQLKELNEAVAALC-------GDQEIMKATE 2382

Qy     172 ----------QVPDEIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDL 220
                    |: : ||:    ::| |   :       |       |  : |  |  ||:|
Db    2383 QSLDPPIEEEHQLRNSIEK----LRARLEADEKKQLCVLQQLKESEHHADLL-KGRVENL 2437

Qy     221 LDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELR-PIKCMEQPSPPSDHSRAYSGKQ 279
           ::   :|        : | : ||:   |:   |:  |: | |     :       :
Db    2438 ERELEIARTN--------QEHAALEAENSKGEVETLKAKIEGMTQSLRGLELDVVTIRSE 2489

Qy     280 KH---DALTRYPQRIRRL--------NIFPANKMEDVQPIDRFVVEEYLLDVLFYLNGCR 328
           |    : | :  :||  |         ||    : | ||       ::|      :  |
Db    2490 KENLTNELQKEQERISELEIINSSFENILQEKEQEKVQ------MKEKSSTAMEMLQTQL 2543

Qy     329 KECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPA-- 386
           ||    :| |           : | ||:  |              ::  : | |   |
Db    2544 KELNERVAALHNDQEACKAKEQNLSSQVECLE------------LEKAQLLQGLDEAKN 2590

Qy     387 ---VVAGAVRALFEKISD---------------LDMESRTRLI-----------LWFSH 416
            |:  :|: | ::: |                 :: : :|:           ||
Db    2591 NYIVLQSSVKGLIQEVEDGKQKLEKKDEEISRLKNQIQDQEQLVSKLSQVEGEHQLWKEQ 2650

Qy     417 HLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEI----------LQREVRLSYWDKIK--Q 464
           :|         |:   ||       |   :|:           |: |: |:  ||:   :
Db    2651 NLELRNLTVELEQKIQVLQ-----SKNASLQDTLEVLQSSYKNLENELELTKMDKMSFVE 2705

Qy     465 SIENATALEELLPPKAGPNFMYSLEEGKEKT---------EEQQLSAELSRKVKEKQTAR 515
           :   || |  |           : |  :||          |: :|: ||   ::| ::::
Db    2706 KVNKMTAKETEL--------QREMHEMAQKTAELQEELSGEKNRLAGELQLLLEEIKSSK 2757

Qy     516 DMIVWIEETIYPVHGFEVTL-TIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLC------ 568
           | :              |:||    ::  ||   |        :|: |:|  ::
Db    2758 DQLK-----------ELTLENSELKKSLDCMHKD------QVEKEGKVREEIAEYQLRL 2799

Qy     569 --PDNDKQVMLLS-------QVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPEN 619
             :   | :||        :: || :              :| | :
Db    2800 HEAEKKHQALLLDTNKQYEVEIQTYRE----------------KLTSKEEC--------- 2834

Qy     620 VDQFHVSDQPWEI---------LGNALNKTYNRISDLRKDISNITK--NVLVAEKASANA 668
                :| |  ||         | |:|  |   : :|:|   :  |  | |  |   |
```

```
Db    2835 -----LSSQKLEIDLLKSSKEELNNSLKATTQILEELKKTKMDNLKYVNQLKKENERAQG 2889

Qy     669 RVEL------------EAAESKLSLVEG----------EPVLGENPAKMKRLKSTV-EK 704
           :::|             |  : :|| ::           :  : |   ::| || |: ||
Db    2890 KMKLLIKSCKQLEEEKEILQKELSQLQAAQEKQKTGTVMDTKVDELTTEIKELKETLEEK 2949

Qy     705 TGEAE---------LSLRESLE-AKEALLNRALSETEVLLLLLFQS--------FLGVLK 746
           | ||:          |   | || ||| |      ||:|  |   ||         || :
Db    2950 TKEADEYLDKYCSLLISHEKLEKAKEML------ETQVAHLCSQQSKQDSRGSPLLGPV- 3002

Qy     747 ERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLTAF 806
             :| |: : || : | : :  :| |
Db    3003 --VPGPSPIPSVTE-KRLSSGQNKASG------------------------------- 3026

Qy     807 TRQYASEIWPH------------MEKLESEVFSGEDVHP 833
            || :| || :             :| :  | ||  :||
Db    3027 KRQRSSGIWENGRGPTPATPESFSKKSKKAVMSG--IHP 3063
```

```
RESULT  15
US-09-356-952-2
; Sequence 2, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
;  APPLICANT: Boriack-Sjodin, Ann
;  APPLICANT: Margarit, S. M.
;  APPLICANT: Bor-Sogi, Dafna
;  APPLICANT: Cole, Philip
;  APPLICANT: Kuriyan, John
;  TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: 600-1-228N
;  CURRENT APPLICATION NUMBER: US/09/356,952
;  CURRENT FILING DATE: 1999-07-19
;  EARLIER APPLICATION NUMBER: 60/093,631
;  EARLIER FILING DATE: 1998-07-21
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1333
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-356-952-2
```

```
  Query Match             2.6%;  Score 112;  DB 3;  Length 1333;
  Best Local Similarity   19.5%;  Pred. No. 0.22;
  Matches  140;  Conservative   97;  Mismatches  200;  Indels  280;  Gaps   38;

Qy     273 RAYSGKQKHDALTRYPQRIRRLNIFPAN----KMEDV-QPIDRFVVEEYLLDV------- 320
           | || : |   |      |:::|    |    : :|: |  : |::|  |  |
Db     406 RFYSQQMKGKQLA-----IKKMNEIQKNIDGWEGKDIGQCCNEFIMEGTLTRVGAKHERH 460

Qy     321 LFYLNGCRKECASY--MANLPVTFRYEYLMAETLFSQILLLPQPPFKTLY---YTLVIMD 375
           :|  :|    | |      ||      || : | | : : :          |   : ::: |
Db     461 IFLFDGLMICCKSNHGQPRLPGASNAEYRLKEKFFMRKVQINDKDDTNEYKHAFEIILKD 520

Qy     376 LCKALPGA---------FPAVVAGAVRALFEKISDLDMESRTRLILWFSHHLSNFQFIWP 426
             :  |           |:::   |:  |:: |: |
Db     521 ENSVIFSAKSAEEKNNWMAALISLQYRSTLERMLDVTM---------------------- 558

Qy     427 WEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIK----QSIENATALEELLPPKAG- 481
                              :|| : ::|| |  :     | ||    || : ||||
Db     559 ------------------LQEEKEEQMRLPSADVYRFAEPDSEEN-IIFEENMQPKAGI 598

Qy     482 ----------------------PNFM----------------------YSLEEGKEKTE 496
                                  |||:                      : : |  | ||
Db     599 PIIKAGTVIKLIERLTYHMYADPNFVRTFLTTYRSFCKPQELLSLIIERFEIPE-PEPTE 657

Qy     497 ---------EQQLSAELSRKVKEKQTARDMIV------WIEETIYPVHGFEVTLTIVVQT 541
                  :| ||||| |  ||       : |      |:|   |    ||
Db     658 ADRIAIENGDQPLSAELKRFRKEYIQPVQLRVLNVCRHWVEHHFY---DFE--------- 705
```

```
Qy     542 LLDIGSKSFTHLVTVLERY-GQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMM 600
                 :|:  :| : | |  |      | |  ::::       :|   :| |
Db     706 -------RDAYLLQRMEEFIGTVRGKAM---KKWVESITKI-------IQRKKIARDNGP 748

Qy     601 GYRLV--SNQAIVRWVFS-PENVDQFHV---------------SD-----QPWEILGNA 636
           |: :   |:   | |  | | ::: | :               ||     || |::|:
Db     749 GHNITFQSSPPTVEWHISRPGHIETFDLLTLHPIEIARQLTLLESDLYRAVQPSELVGSV 808

Qy     637 LNKTYNRIS--DLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEG---------- 684
             |    |:  :| | | : |   |  ||     || |  | ::::|
Db     809 WTKEDKEINSPNLLKMIRHTTNLTLWFEKCI----VETENLEERVAVVSRIIEILQVFQE 864

Qy     685 --------EPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLNRALSETEVLLLL 736
                 | |   | : : ||  | |:    :  : |  || |      |||
Db     865 LNNFNGVLEVVSAMNSSPVYRLDHTFEQIPSRQKKILE--EAHE------LSEDH----- 911

Qy     737 LFQSFLGVLKERLPDPTKVRSVQD---------LKSIGAEDDKPSAMDVDSENGNP---- 783
            :: :|           |:||:           | :|           : :| |||
Db     912 -YKKYL----------AKLRSINPPCVPFFGIYLTNI----------LKTEEGNPEVLK 949

Qy     784 ---KKSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEV---FSGEDVHPL 834
             |:     :| :     :  :|   :|| ::  |:  ::||::   |  |:::|:
Db     950 RHGKELINFSKRRK-----VAEITGEIQQYQNQ--PYCLRVESDIKRFF--ENLNPM 997
```

Search completed: August 20, 2002, 15:26:30
Job time: 5254 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:51:51 ; Search time 63.71 Seconds
(without alignments)
1478.427 Million cell updates/sec

Title: US-09-882-986-2
Perfect score: 4374
Sequence: 1 MSNWKTLLLRIGEKGPEYGT.....EDVHPLFLQAISSALQFPLH 848

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	818.5	18.7	800	22	ABB62896	Drosophila melanog
2	550.5	12.6	747	22	ABB63581	Drosophila melanog
3	145.5	3.3	1833	22	ABB71141	Drosophila melanog
4	141.5	3.2	5373	22	AAU14603	Novel bone marrow
5	140	3.2	906	22	AAG67251	Amino acid sequenc
6	140	3.2	906	22	AAB94323	Human protein sequ
7	138.5	3.2	3868	22	ABB63663	Drosophila melanog
8	137.5	3.1	814	22	AAM93442	Human polypeptide,
9	135	3.1	3854	22	ABB67104	Drosophila melanog
10	135	3.1	5385	22	ABB66487	Drosophila melanog
11	135	3.1	5496	22	ABB67161	Drosophila melanog
12	135	3.1	8805	22	ABB67112	Drosophila melanog
13	134.5	3.1	1084	22	AAB93096	Human protein sequ
14	133.5	3.1	710	22	AAM78497	Human protein SEQ
15	133.5	3.1	3263	22	ABB67210	Drosophila melanog
16	132	3.0	1743	19	AAW98879	H. pylori GHPO 175
17	131.5	3.0	2154	20	AAW81639	Mouse elf-1 protei
18	131	3.0	2633	22	ABG06505	Novel human diagno
19	131	3.0	2688	22	AAM40883	Human polypeptide
20	131	3.0	5447	22	AAU14697	Novel bone marrow
21	130	3.0	6815	22	ABB66811	Drosophila melanog
22	129.5	3.0	981	22	AAU36618	Staphylococcus aur
23	129.5	3.0	2272	18	AAW21731	GAL4/HA/NuMA fusio
24	129	2.9	2663	22	AAM39097	Human polypeptide
25	128	2.9	4590	22	AAU33184	Novel human secret
26	128	2.9	7201	22	ABB71136	Drosophila melanog
27	127	2.9	980	22	AAB62027	Recombinant P. fur
28	127	2.9	1003	20	AAY34599	Chlamydia pneumoni
29	127	2.9	4536	19	AAW41262	Apolipoprotein B-1
30	126.5	2.9	808	22	ABG05140	Novel human diagno
31	126.5	2.9	1269	22	AAM40297	Human polypeptide
32	126.5	2.9	2013	22	ABB62322	Drosophila melanog
33	125.5	2.9	2415	22	ABG20279	Novel human diagno
34	125	2.9	469	19	AAW37881	BRCA1 modulator pr
35	125	2.9	469	20	AAY30149	Amino acid sequenc
36	125	2.9	1671	11	AAR06341	Signal peptide and
37	125	2.9	1829	18	AAW29322	DNA polymerase wit
38	124.5	2.8	2400	22	ABG20278	Novel human diagno
39	123	2.8	677	20	AAW85608	Secreted protein c
40	123	2.8	677	22	AAG64560	Human myosin heavy
41	123	2.8	809	20	AAY09371	Human apolipoprote
42	123	2.8	1303	22	ABB62088	Drosophila melanog
43	123	2.8	2721	8	AAP70647	Sequence of N-term
44	123	2.8	4536	20	AAW96826	Amino acid sequenc
45	122.5	2.8	1521	21	AAG39235	Arabidopsis thalia

ALIGNMENTS

RESULT 1
ABB62896
ID ABB62896 standard; Protein; 800 AA.
XX
AC ABB62896;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 15480.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL06999.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

```
XX
PS   Disclosure; SEQ ID NO 15480; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   800 AA;


  Query Match             18.7%;  Score 818.5;  DB 22;  Length 800;
  Best Local Similarity   25.5%;  Pred. No. 8.6e-61;
  Matches  217;  Conservative  178;  Mismatches  340;  Indels  115;  Gaps   19;

Qy       5 KTLLLRIGEKGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHKIPLY 64
Db      35 eslilrvger-----stssvesnleglvsvleadlgtfrlkilrilsdcavrmpekctvy 89

Qy      65 GTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLI 124
Db      90 ttlvgllnaknykfggefvdhmvktfkeslkmcrwdaaryslrfladlvnchvisatsll 149

Qy     125 VVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGI 184
Db     150 qlldtmid-----vsnedtvp--qvrrdwfvfavlstlpwvgrdlyekkesalesllrri 202

Qy     185 QAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSF 244
Db     203 evylnkrskkhhnalrvwss---daphpqeeyldclwaqirklrqdnwaekhiprpylvf 259

Qy     245 EAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMED 304
Db     260 dsilcealqhnlpti------vppphhdn----------feypmpwvvyrmfdytdcpd 302

Qy     305 ------VQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFR--YEYLMAETLFSQI 356
Db     303 gpnlpgahsierflieehlhhiietyhherkdcaaqllsfpykhkipleycivevvfael 362

Qy     357 LLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSH 416
Db     363 fhmptpryldicygsilielcklqpatlpqvlaqateilfmridsmntscfdrfvnwfsy 422

Qy     417 HLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELL 476
Db     423 hlsnfkftwswdewdsclslldgehprpkfiqevlqkclrlsyhqritemm--pttyakli 480

Qy     477 PPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFE---- 532
Db     481 pltpvpnykyaneea-anlpgttvahqlvvairqkctpeevvnilkd--ipnsgysgeem 537

Qy     533 -------VTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYW 585
Db     538 sdgsfnalkidvfvqtllnlgsksfshsfaaiskfhsvfralaeteeaqicilhnifelw 597

Qy     586 KNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNRIS 645
Db     598 sshqqmmvvlidkllklqivdcsavatwifskemtgef-tklylweilhltikkmnkhvi 656

Qy     646 DLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKT 705
Db     657 klnteiseakeklakadssssds--------------------eddsshkrkkpit--- 692

Qy     706 GEAELSLRESLEAKEALLNRALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIG 765
Db     693 -hadkpseevvermeekleaanvnqkrlflivfqrfimilsehl-----lrsdtd----- 741
```

```
Qy     766 AEDDKPSAMDVDSENGNPKKSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEV 825
Db     742 -------grdpdtd--------------wyrwtigrlqqvflmhheqvqkysstletll 779

Qy     826 FSGE-DVHPL 834
Db     780 ftsdldthil 789


RESULT   2
ABB63581
ID   ABB63581 standard; Protein; 747 AA.
XX
AC   ABB63581;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 17535.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07684.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 17535; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   747 AA;


  Query Match             12.6%;  Score 550.5;  DB 22;  Length 747;
  Best Local Similarity   25.0%;  Pred. No. 7.2e-38;
  Matches  167;  Conservative  136;  Mismatches  257;  Indels  107;  Gaps   22;

Qy       7 LLLRIGEKGPEY----GTSSDYKDHIETCFGVIRREIERSGDQVLPFLLQCAEQLPHKIP 62
Db      45 lelklgnlg--yllgkgmsvdlkenl----------------inelviclgyypgqas 84

Qy      63 LYGTLIGLLNLENEDF--------VQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLC 114
Db      85 ayatlvglinvadfefgsecllfmahkvgesvhir--------dwnrcrgvvhcmvdlyh 136
```

```
Qy    115 SKVIQPASLIVVFETLLSSAATTVD-EEKGNPSWQPQADFYVICILSSLPWGGSELAEQV 173
Db    137 cqvlpssyilkllaaflkdcealkdpddlvgvtpqirrdflaycvlsampligrdleget 196

Qy    174 PDEIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGW- 232
Db    197 --afdklivslqiyikkrsalhtnmlsvw------pdfnqrdylellwqqvdgmrqnhwa 248

Qy    233 --KLESVPRPHLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQR 290
Db    249 epehqliprpyksfsetlshgrihqlr-----------dydla-----aheercryplp 291

Qy    291 IRRLNIFPANKMEDVQ------PIDRFVVEEYLLDVLFYLNGCRKECAS----YMANLPV 340
Db    292 rvcfrifscdsvgeipnmpppvsierhlleahildilisfhkerkicadsllmyaaskpq 351

Qy    341 TFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKIS 400
Db    352 lpvy-ycivevilgemlrlptanwstiaygsilvelckrqpdkipqvvaqaldiiynrln 410

Qy    401 DLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLP------KWAPKRVFVQEILQREV 454
Db    411 smsvacfdrlvnwvshhisnfgfncqwskwaqglpspidpsatnlqpkvvflrellkkcf 470

Qy    455 RLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKVKEKQTA 514
Db    471 r------ikdvvpd--iladflpppvlphfkfv----detlpgailskdlleamrspqas 518

Qy    515 RDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKL-CPDNDK 573
Db    519 pemiseiiksstgi-gpllkinvftqnclhlgsksfshtfgilrkyhsvfkdlvagdper 577

Qy    574 QVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSP---ENVDQFHVSDQPW 630
Db    578 qaavlngifdvwvasdqykfvvteklvtgliiepisvvrwifgpsmrkeltkiyi----w 633

Qy    631 EILGNAL 637
Db    634 ellhsal 640


RESULT   3
ABB71141
ID   ABB71141 standard; Protein; 1833 AA.
XX
AC   ABB71141;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 40215.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15244.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 40215; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1833 AA;


  Query Match             3.3%;  Score 145.5;  DB 22;  Length 1833;
  Best Local Similarity   19.7%;  Pred. No. 0.017;
  Matches  177;  Conservative  112;  Mismatches  319;  Indels  291;  Gaps   32;

Qy    118 IQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPD-- 175
Db    235 lppkkvavastttasssnsssts1ktsnstsasnevkvvtstsssstsssvrrkeadsv 294

Qy    176 ---EIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGW 232
Db    295 askeikrqtv---paasishsnstss---------tastask----------sqdtngm 331

Qy    233 KLESVPRPHLSFEAQLVAGKFHE-------LRPIKCMEQPSPPSDHSRAYSGKQK----H 281
Db    332 q-eqmkalkleletmktraekaereksdillrrlasmdtasnrtaasealnlqqklnemk 390

Qy    282 DALTRYPQRIRRLNIFP---ANKMEDVQPIDRFVVEEYLLDVLFYLN-GCRKECASYMAN 337
Db    391 eqldrvtedkrklnlrmkelenkgseselrrklqaaeqiceelmeenqsakkeilnlqae 450

Qy    338 LPV---TFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRA 394
Db    451 mdevqdtfrddevkaktslqk------------------dlekatknc--rilsfklkk 489

Qy    395 LFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEILQREV 454
Db    490 sdrkietleqerqss----fnaelsn--------------------------kikklee 518

Qy    455 RLSYWDKIKQSIENATALEELLPP--KAGPNF--------------MYSLEEGKEKTEEQ 498
Db    519 elrfsneltrklq--aeaeelrnpgkkkapmlgvlgkstsadakftresltrggsqedpq 576

Qy    499 QLSAELSRKVKEKQTARDMIVWIEETIYPVHG-------------FEVTLTIV------- 538
Db    577 hlqrelqdsieretdlkdqlkfaeeelqrlrdrerkrvrfscgtqtevplevvafprgtq 636

Qy    539 -VQTLLDIGSKSFTHLVT---------------------------------VLERYGQ 562
Db    637 tvatvqsdmstsvenlvtsnvavtqtdfevpdrnvsieretmsspfaglfppssssrvgq 696

Qy    563 VFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVF----SPE 618
Db    697 sgrklsptphphrl-------------apevhadrdegisdeddpaelrillelneqea 742

Qy    619 NVDQFHVSDQPWEILGNALNKTYNR--ISDLRKDISNITKNVLVAEKASANARVE----- 671
Db    743 silrlkvedleke---naeskkyvrelqaklrqdssngsksslllslgtsssaaekkvktl 799

Qy    672 -------------------LEAAESKLSLVEGE-PVLGENPAKMKRLKSTVEKTGEAEL 710
Db    800 neelvqlrrtltekeqtvdslknqlskldtletendklakenkrllalrkasektgevdq 859
```

```
Qy    711 SLRES------------------LEAKEAL----------------------LNRALS 728
          ::||                  |||:: |                        |   :|
Db    860 kmkeslaqaqrerdeltarlkrmqleaedklpprtakrvndltpkshlkkwveeledeis 919

Qy    729 ETEVLL-------LLLFQSFLGVLKERLPDPTKVRS-----VQDLKSIGAEDDKPSAMDV 776
          |  |:|       |   ||  | |:| |    :  |     || || :     | | ::
Db    920 emrvmlsssgtdqlkalqsakgaleedlrkckqklslaegdvqrlkllngssskvseleq 979

Qy    777 DSENGNPK--------KSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEVFS 827
            : |: :         |  |   ::|     ||       : :   |    | | |  |
Db    980 klkrgdeeakklnsklkdledkvkkqeaqlklg-------etskstwesqskrekekls 1031


RESULT   4
AAU14603
ID   AAU14603 standard; Protein; 5373 AA.
XX
AC   AAU14603;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Novel bone marrow polypeptide #2.
XX
KW   Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW   haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW   wound healing; nutritional supplement; immune disorder;
KW   severe combined immunodeficiency; SCID.
XX
OS   Homo sapiens.
XX
PN   WO200157187-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03782.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   30-NOV-2000; 2000US-0250683.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Ford JE,  Boyle BJ,  Tang YT,  Liu C,  Asundi V,  Zhou P,  Xue AJ;
PI   Ren F,  Drmanac RT;
XX
DR   WPI; 2001-488875/53.
DR   N-PSDB; AAS22908.
XX
PT   Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT   and gene therapy -
XX
PS   Claim 10; Page 198-208; 392pp; English.
XX
CC   AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC   invention. The proteins and corresponding coding sequences may be used
CC   in the prevention, diagnosis and treatment of diseases associated with
CC   inappropriate bone marrow polypeptide expression. For example, to treat
CC   disorders associated with decreased expression by rectifying mutations
CC   or deletions in a patient's genome that affect the activity of the
CC   polypeptides by expressing inactive proteins or to supplement the
CC   patient's own production of the polypeptide. Additionally, the nucleic
CC   acids may be used to produce the polypeptides, by inserting the nucleic
CC   acids into a host cell and culturing the cell to express the protein.
CC   The nucleic acid and its complementary sequences may also be used as DNA
CC   probes in diagnostic assays to detect and quantitate the presence of
CC   similar nucleic acid sequences in samples, and therefore which patients
CC   may be in need of restorative therapy. The proteins may also be used as
CC   antigens in the production of antibodies against bone marrow proteins
CC   and in assays to identify modulators of their expression and activity.
CC   The anti-bone marrow protein antibodies and antagonists may also be used
CC   to down regulate expression and activity. The antibodies may also be used
CC   as diagnostic agents for detecting the presence of the protein in samples
CC   (e.g. by enzyme linked immunosorbant assay (ELISA)). The proteins
CC   may be used to regulate haematopoiesis activity, and consequently in the
CC   treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC   such as wound healing; as a nutritional supplement;  and in treatment of
CC   immune disorders such as severe combined immunodeficiency (SCID).
XX
SQ   Sequence   5373 AA;


  Query Match             3.2%;  Score 141.5;  DB 22;  Length 5373;
  Best Local Similarity   19.7%;  Pred. No. 0.22;
  Matches  161;  Conservative  133;  Mismatches  341;  Indels  183;  Gaps   35;

Qy     52 QCAEQL-PHKIP---LYGTLIGLLNLENED-----FVQKLVESVHANFQVALDSGNCNSI 102
          |  |:|  |::|   |      :: :| |       ||:  :|: ::||         |:
Db   1944 qlqeelaehqvpveklqkvardimeiegepapdhrhvqettdsilshfq---------sl 1994

Qy    103 RILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSL 162
            |   :|||   : |  |:    |:|| |        :|                :|||
Db   1995 syslaersslqkaiaqsqsvqeslesllqsigeveqnlegkq------------vssl 2041

Qy    163 PWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLD 222
            | : |     :::: :   :: |   :   |    |    :  ::   :  : ::
Db   2042 ssgviqkalatnmklkqdiarqkssleatremvt---rfmetadsttaavlqgklaevsq 2098

Qy    223 RIQSLASNGWKLES-----VPRPHLSFEAQLVAGKFHELRPIKCMEQPSPPSDHSRAYSG 277
          | : |     : ||     :|:  : ||   ::||  :      ||  |         ||
Db   2099 rfeqlclqqqekesslkkllpqaem-feh--lsgklqqf-----menks-----rmlasg 2145

Qy    278 KQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVE----EYLLDVLFY---LNGCRKE 330
            |    :| : |:|: ||:   :: |::  ::  | |      : ||:  :   :   ||:
Db   2146 nqpdqdithffqqiqelnlemedqqenldtlehlvtelsscgfaldlcqhqdrvqnlrkd 2205

Qy    331 CASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKALPGAFPAVVAG 390
                  |  | :      | :   |:       |: |  |        |   |: |
Db   2206 ----ftelqktvkerekdasscqeql-----defrklvrt--fqkwlketegsipptet- 2253

Qy    391 AVRALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEIL 450
                   :|  ::| :          || :                |  || |   |:||
Db   2254 -------svsakelekqi-------ehlksl--------------lddwaskgtlveein 2285

Qy    451 QREVRLS--YWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELSRKV 508
            :   |     :     :  | | |:   :  |  |     |      :: | | |:|
Db   2286 ckgtslenlimeitapdsqgktdlteiqcdmsdvnlkyeklggvlherqeslqailnrme 2345

Qy    509 KEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHLVTVLERYGQVFSKLC 568
            :  : |  :: |:|          :   |:::  : | : |  |       : |
Db   2346 evhkeansvlqwle----------skeevlksmdamssptktetvkaqaesnkaflael 2394

Qy    569 PDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQ 628
            |  ::  : :           |:|             :| :  | |         :
Db   2395 eqnspkiqkvke----------ala----------gllvtypnsqe--------ae 2422

Qy    629 PWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLS--LVEGEP 686
            |: :   ||  : | :::      : :   : | ||  |    :||||:|   |:| |
Db   2423 nwkkiqeelnsrweratev-----tvarqrqleesashlac--fqaaesqlrpwlmekel 2475

Qy    687 VLGE------NPAKMKRLKSTVE---KTGEAELSLRESL-EAKEALL----NRALSETEV 732
          ::|       :|  :   |  |:   |  ||     | | || : :|    : :|| ::|
Db   2476 mmgvlgplsidpnmlnaqkqqvqfmlkefearrqqheqlneaaqgiltgpgdvslstsqv 2535

Qy    733 LLLLLFQSFLGVLKER---LPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCEV 789
              |  | : ::   | |    || |  ::|          :   ||      |  |
Db   2536 ------qkelqsinqkwveltdklnsrssqidqaivkstqyqellqdlse-----kvrav 2584

Qy    790 GER--EQWCLSTLGYLTAFTRQYASEIWPHMEKLESEV 825
            |:|   |  :||         :  |||   :|:|: ||
Db   2585 gqrlsvqsaistqpeavkqqleetseirsdleqldhev 2622


RESULT   5
AAG67251
ID   AAG67251 standard; Protein; 906 AA.
```

```
XX
AC   AAG67251;
XX
DT   13-NOV-2001  (first entry)
XX
DE   Amino acid sequence of a human liver-associated gene.
XX
KW   Liver-associated gene; hepatic cancer; liver cancer; marker.
XX
OS   Homo sapiens.
XX
PN   WO200109318-A1.
XX
PD   08-FEB-2001.
XX
PF   28-JUL-2000; 2000WO-JP05064.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   18-OCT-1999;   99US-0159590.
PR   11-JAN-2000; 2000JP-0118776.
PR   17-FEB-2000; 2000US-0183322.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T,  Aburatani H;
PI   Kodama T,  Hippo Y,  Taniguchi H;
XX
DR   WPI; 2001-541221/60.
DR   N-PSDB; AAH77804.
XX
PT   New genes and their proteins found in hepatic cancer, vectors,
PT   antibodies, and screening method for compounds with binding activity,
PT   useful as diagnostics and targets for treatment agents -
XX
PS   Claim 1; Page 63-66; 99pp; Japanese.
XX
CC   The present sequence is encoded by a liver-associated gene. The
CC   expression level of this gene changes in liver (hepatic) cancer.
CC   Liver-associated gene can be used as markers in blood tests for
CC   screening for the early stages of the liver cancer. The encoded
CC   proteins and peptides can also be used as targets for screening
CC   for compounds to treat the disease. They can also be used for
CC   predicting the progress of the disease.
XX
SQ   Sequence   906 AA;
```

```
  Query Match             3.2%;  Score 140;  DB 22;  Length 906;
  Best Local Similarity   17.9%;  Pred. No. 0.015;
  Matches  162;  Conservative  144;  Mismatches  307;  Indels  292;  Gaps   37;

Qy      34 VIRREIERSGDQVLPFLLQCAEQLPHKIPLYGTLI--------------GLLNLENEDF 78
           |: | : |:   :  ||      |   | :|  |:                 | |:
Db      82 vlnrvvtrasspishfl-----dllsnivmyaplvlqscsskvteafdylsflplqt--- 133

Qy      79 VQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTV 138
           ||:|:::|    :|::   :|    ::|    ::  :::    | :  |  ||
Db     134 vqrllkavqpllkvsmsmrdc----lilvlrkamfanqldarksavagfllllk------ 183

Qy     139 DEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSG 198
                     :| |:  |||     |: :: :   : :| | : ::
Db     184 ------------nfkvlgslss-----sqcsqsl--svsqvhvdvhshy---------- 213

Qy     199 LNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELR- 257
                  :|:| : |  :::| ::   |                 |  | :  ||
Db     214 ----------nsvanetfcleimdslrrclsq----------qadvrlmlyegfydvlrr 253

Qy     258 --------------PIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKME 303
                         :|   :| |           | |   | ::    :   :|:
```

```
Db     254 nsqlansvmqtllsqlkqfyepkp-------------dllp--plkleaciltqgdkis 297

Qy     304 DVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFR------------YEYL--MA 349
             :|:|              ||  | : | ::  |  :  :            || |  :
Db     298 lqepld-------------yllcciqhclawyknтviplqqgeeeeeeeeafyedlddil 344

Qy     350 ETLFSQILLLPQPPF----------------KTLYYTLVIMDLCKAL-----------PG 382
           |:: ::::     |                 |      ::| :|: |
Db     345 esitnrmikseledfeldksadfsqstsigiknnisaflvmgvcevlieynfsissfskn 404

Qy     383 AFPAVVAGAVRALFEKISDLDME----SRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPK 438
            |  ::  ::   ::|:||:  |    ::|::    |  | : :
Db     405 rfedil--slfmcykklsdilnekagkaktkmanktsdsllsmk---------------- 446

Qy     439 WAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEK---- 494
                 ||  :|    |:  |    :     |::||    : | :      || :|
Db     447 ------fvsslltalfrvllw----rytsiptsveesgkkekgksislicleglqkifsa 496

Qy     495 ------TEEQQLSAELSRKVKEKQTARDMIVWI-EETIYPVHGFEVT-LTIVVQTLLDIG 546
                 : ||    |    || :   |  | : : | : :  |: : | ::     |
Db     497 vqqfyqpkiqqflraldvtdkegeeredadvsvtqrtafqirqfqrsllnllssqeedfn 556

Qy     547 SKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVS 606
           ||    ||||    ::   | | : : | :||  |   | | :  |:    :|
Db     557 skealllvtvltslskl---lepsspqfvqmlswtskickensredalfckslmnl---- 609

Qy     607 NQAIVRWVFSPENVDQFHVS-DQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKAS 665
                 :||       |||   |  :| :     :  : |: :|:          || :
Db     610 -------lfs------lhvsykspvillrdlsqdihghlgdidqdve--------vektn 648

Qy     666 ANARVELEAAESKLSLV---EGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEAL 722
             | | |  |   : |:   : | || |    : :||  |     :: :| |   ::  |
Db     649 hfaivnlrtaaptvcllvlsqaekvleevdwlitklkgqv-----sqetlseeassqatl 703

Qy     723 LNRALSETEVLLL-LLFQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENG 781
           |: : :  :: | |   |   :::  ||  : | ::  || :
Db     704 pnqpvekaiimqlgtlltffhelvqtalpsgscvdtl--lkdl---------------- 744

Qy     782 NPKKSCEVGEREQWCLSTLGYLTAFTRQY------ASEIWPHMEKLESEVFSGEDVHPLF 835
                |::             |||  | |      :  | :||||     ||  : ||
Db     745 -----ckm----------yttltalvryylqvcqssggipknmeklvk--lsgshltplc 787

Qy     836 LQAIS 840
              ||
Db     788 ysfis 792
```

```
RESULT   6
AAB94323
ID   AAB94323 standard; Protein; 906 AA.
XX
AC   AAB94323;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:14805.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
```

```
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14805; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   906 AA;


  Query Match             3.2%;  Score 140;  DB 22;  Length 906;
  Best Local Similarity   17.9%;  Pred. No. 0.015;
  Matches  162;  Conservative  144;  Mismatches  307;  Indels  292;  Gaps   37;

Qy      34 VIRREIERSGDQVLPFLLQCAEQLPHKIPLYGTLI--------------GLLNLENEDF 78
           |: | : |:   :  ||      |   | :|  |:                 | |:
Db      82 vlnrvvtrasppishfl-----dllsnivmyaplvlqscsskvteafdylsflplqt--- 133

Qy      79 VQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTV 138
           ||:|:::|    :|::   :|    ::|    ::  :::    | :  |  ||
Db     134 vqrllkavqpllkvsmsmrdc----lilvlrkamfanqldarksavagfllllk------ 183

Qy     139 DEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSG 198
                     :| |:  |||     |: :: :   : :| | : ::
Db     184 ------------nfkvlgslss-----sqcsqsl--svsqvhvdvhshy---------- 213

Qy     199 LNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELR- 257
                   :|:| : |  :::| ::   |                 |  | :  ||
Db     214 ----------nsvanetfcleimdslrrclsq----------qadvrlmlyegfydvlrr 253

Qy     258 --------------PIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKME 303
                        :|   :| |              | |   | ::    :   :|:
Db     254 nsqlansvmqtllsqlkqfyepkp-------------dllp--plkleaciltqgdkis 297

Qy     304 DVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFR------------YEYL--MA 349
             :|:|             ||  | : | ::  |  :  :             || |  :
Db     298 lqepld------------yllcciqhclawyktviplqqgeeeeeeeeafyedlddil 344

Qy     350 ETLFSQILLLPQPPF---------------KTLYYTLVIMDLCKAL-----------PG 382
           |:: ::::     |                |      ::| :|: |
Db     345 esitnrmikseledfeldksadfsqstsigiknnisaflvmgvcevlieynfsissfskn 404

Qy     383 AFPAVVAGAVRALFEKISDLDME----SRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPK 438
                 | ::  ::   ::|:||:  |    ::|::    |  | : :
Db     405 rfedil--slfmcykklsdilnekagkaktkmanktsdsllsmk--------------- 446

Qy     439 WAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEK---- 494
                 ||  :|    |:  |    :     |::||    : | :      || :|
Db     447 -----fvssllталfrvllw----rytsiptsveesgkkekgksisllcleglqkifsa 496

Qy     495 ------TEEQQLSAELSRKVKEKQTARDMIVWI-EETIYPVHGFEVT-LTIVVQTLLDIG 546
                  : ||    |    || :   |  | : : | : :  |: : | ::     |
Db     497 vqqfyqpkiqqflraldvtdkegeeredadvsvtqrtafqirqfqrsllnllssqeedfn 556

Qy     547 SKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVS 606
           ||    |||||    ::   | | : : | :||  |   | | :  |:    :|
Db     557 skealllvtvltslskl---lepsspqfvqmlswtskickensredalfckslmnl---- 609

Qy     607 NQAIVRWVFSPENVDQFHVS-DQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKAS 665
                  :||       |||   |  :| :     :  : |: :|:          || :
Db     610 -------lfs------lhvsykspvillrdlsqdihghlgdidqdve--------vektn 648

Qy     666 ANARVELEAAESKLSLV---EGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEAL 722
             | | |  |   : |:   : | || |    : :||  |     :: :| |   ::  |
Db     649 hfaivnlrtaaptvcllvlsqaekvleevdwlitklkgqv-----sqetlseeassqatl 703

Qy     723 LNRALSETEVLLL-LLFQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENG 781
            |: : :  :: | |   |  ::: || : | ::  || :
Db     704 pnqpvekaiimqlgtlltffhelvqtalpsgscvdtl--lkdl---------------- 744

Qy     782 NPKKSCEVGEREQWCLSTLGYLTAFTRQY------ASEIWPHMEKLESEVFSGEDVHPLF 835
                 |::              |||  | |      :  |  :|||     ||  : ||
Db     745 -----ckm----------yttltalvryylqvcqssggipknmeklvk--lsgshltplc 787

Qy     836 LQAIS 840
             ||
Db     788 ysfis 792


RESULT   7
ABB63663
ID   ABB63663 standard; Protein; 3868 AA.
XX
AC   ABB63663;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 17781.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07766.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 17781; 21pp + Sequence Listing; English.
XX
```

```
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   3868 AA;

  Query Match             3.2%;  Score 138.5;  DB 22;  Length 3868;
  Best Local Similarity   17.6%;  Pred. No. 0.23;
  Matches  174;  Conservative  148;  Mismatches  342;  Indels  323;  Gaps   40;

Qy       8 LLRIGEKGPEYGTSSDYKDH---IETCFGVIRREIERSGDQVLPFLLQCAEQLPHKIPLY 64
Db     315 mmrmslewsefssdldtrpnawsilrtfaeiateismvgyrmep--lqpqvqtltsmaay 372

Qy      65 GTLIGLLNLE-NEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASL 123
Db     373 akvndylkiemndnflkdvidkvq--------------niilr---------------- 401

Qy     124 IVVFETLLSSAATTVDEEKGNPSWQ---------------------------------- 148
Db     402 --tyqeviqyvegfrdkyyalyswqerdalnqflsephefeeyfaridmyygfiqmlrse 459

Qy     149 PQADFYVICILSSLP--WGGSELAEQVPDEIERVLV--GIQA------------YLSIRK 192
Db     460 pateyfvmavihnepaifglrtlaenliheittiiirehikaevdicdefekikyralei 519

Qy     193 NSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGK 252
Db     520 pksteel--lesaeymihv-kkdkiaeltdriqyclqvgtnivelte----------msk 566

Qy     253 FHELRPIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKMEDVQPIDRFV 312
Db     567 yhfdltikti----------------------nwikdindicdynasqqeqf----kft 599

Qy     313 VEEYLLDVLFYLNGCRKECASYMA-----NLPVTFRYEYLMAETLFSQI---------LL 358
Db     600 feehlqevikklnsdidellpkltviddmsrpdkfrdsyiilqnfidqlktfddyvawin 659

Qy     359 LPQPPFKTLYYTLVIMDLCKALPGAFPAVVAGAVRALFEKISDLDMESRTRLILWFSHHL 418
Db     660 keeklfkvalteypkldiiktf--vyp----------faelmkcciewqrylsvw---nd 704

Qy     419 SNFQFIWPWEEWAFVLDLPKWAPK-----RVFVQEIL-----------------QREVR 455
Db     705 gpfeylepqfverttddylkefqknqkyyrvkikqdlidnpvckfkgqtedpdpakhpvp 764

Qy     456 LSYWDKIKQSIENATA-------------------LEEL---LPPKAGPNFMYSLEEG 491
Db     765 lrlctsmiqsikdfttgvfivntmcnpalrkrhwkemseiagfdvtpdagttlrkilnsg 824

Qy     492 KE------------KTEEQQLSAELSRKVKEKQTARDMIVWIEETIYPVHGFEVTLTIVV 539
Db     825 ldpildqfeiisigankelqlwnalqamike---------w-etrvfpygpyketgvqil 874

Qy     540 QTLLDIGSKSFTHLVTVLERYGQVFSKLCPDNDK----QVMLLSQVSTYWKNNVQMTAVA 595
Db     875 sslddiqallddhilktlvmrgsafmkpceeevrawyekimrvnetldqwg--------- 925

Qy     596 IDRMMGYRLVSNQAIVRWVFSPENVDQFHVSDQPWE-ILGNALNKTYNRISDLRKDISNI 654
Db     926 -------kvqanylyllpifsskdi----vaqmpeegrlfviveqtytr------nmglv 968

Qy     655 TKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELSLRE 714
```

```
Db     969 lrqplvmetapvsgl--leslqkanellediatgvsnylekkrlyfp----------rf 1015

Qy     715 SLEAKEALLNRALSETEVLLLLL-------------FQSFLGVLKERLPDPTKVRSVQD 760
Db    1016 fflandeml-eilsetkdplrvlphlskcfeginslefdaaknvlamissdketiefieq 1074

Qy     761 LKS----------IGAEDDKPSAMDVDSENG---NPKKSCEVGEREQWCLSTLGYLTAF 806
Db    1075 vstaaaggsvekwligvedemlkavryqnelsfahypkv-----krhewvle-------- 1121

Qy     807 TRQYASEIWPHMEKLE-SEVFSGEDVH 832
Db    1122 --------wpqmtvlaisqvywasrvh 1140


RESULT   8
AAM93442
ID   AAM93442 standard; Protein; 814 AA.
XX
AC   AAM93442;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide, SEQ ID NO: 3084.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-0114089.
XX
PR   08-JUL-1999;   99JP-0194486.
PR   11-JAN-2000; 2000JP-0118774.
PR   02-MAY-2000; 2000JP-0183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2001-524255/58.
DR   N-PSDB; AAK94363.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their
PT   use in genetic manipulation -
XX
PS   Claim 8; SEQ ID NO 3084; 1380pp + sequence listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been
CC   isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC   molecules have been determined. Primers for synthesising the full length
CC   cDNA are useful for clarifying the function of the protein encoded by
CC   the cDNA. The full length clones were obtained by construction of full
CC   length enriched cDNA libraries that were synthesised by the oligo-capping
CC   method. The primers enable the production of the full length cDNA easily
CC   without any special methods. The present sequence is a polypeptide
CC   encoded by a full length human cDNA of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ   Sequence   814 AA;

  Query Match             3.1%;  Score 137.5;  DB 22;  Length 814;
  Best Local Similarity   19.9%;  Pred. No. 0.021;
  Matches  108;  Conservative   92;  Mismatches  244;  Indels   99;  Gaps   24;

Qy     341 TFRYEYLMAETLFSQILLLPQP--PFKTLYYTL--VIMDLCKALPG-AFPAVVAG----- 390
```

```
Db    277 sirysgdssktvlmalclmalpfipasnlffpvgfvvaervlyvpsmgfcilvahgwqki 336

Qy    391 AVRALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKRVFVQEIL 450
Db    337 stksvfkklswicls-----mvilthslktfhrnwdwe----------seytlfmsalk 380

Qy    451 QREVRLSYWDKIKQSIENATALEELL----------PPKAGPNF-----MYSLEEGKEKT 495
Db    381 vnknnaklwnnvghaleneknferalkyflqathvqpddigahmnvgrtyknlnrtkeae 440

Qy    496 EEQQLSAELSRKV--KEKQTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDIGSKSFTHL 553
Db    441 esymmakslmpqiipgkkyaar---------iapnh-----lnvyi-nlanliranesrl 485

Qy    554 VTVLERYGQVFSKLCPDNDKQV------MLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSN 607
Db    486 eeadqlyrqais-mrpd-fkqayisrgelllkmnkplkakeaylkaleldr--------n 535

Qy    608 QAIVRWVFSPENVDQFHVS-DQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASA 666
Db    536 nadl-wy----nlaivhielkepneal-----knfnralelnpkhklalfnsaivmqesg 585

Qy    667 NARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELSLRESL----EAKEAL 722
Db    586 evklrpearkrllsyineepldangyfnlgmlamddkkdneaeiwmkkaiklqadfrsal 645

Qy    723 LNRALSETEVLLLLLFQSFLGVLKERLPDPTK-------VRSVQDLKSIGAEDDKPSAMD 775
Db    646 fnlallysqtakelkalpileellryypdhikglilkgdilmnqkkdilgakkcferile 705

Qy    776 VDSENGNPKKS-CEVGEREQWCLSTLGYL--TAFTRQYASEIWPHMEKLESEVFSGEDVH 832
Db    706 mdpsnvqgkhnlcvvyfeekdllkaercl1etlalapheeyiqrhlnivrdkisssfie 765

Qy    833 PLF 835
Db    766 pif 768


RESULT   9
ABB67104
ID   ABB67104 standard; Protein; 3854 AA.
XX
AC   ABB67104;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 28104.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL11207.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 28104; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   3854 AA;


  Query Match           3.1%;  Score 135;  DB 22;  Length 3854;
  Best Local Similarity   19.1%;  Pred. No. 0.46;
  Matches  183;  Conservative  156;  Mismatches  348;  Indels  270;  Gaps   43;

Qy     55 EQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCN--SIRILLRFMTSL 112
Db   2186 erllqsiaehkplldklnktgealgalvadddgakineildtdnaryaalrlelrerqqa 2245

Qy    113 LCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQ 172
Db   2246 lesalqessqfsdklegmlralantvdqvn-------qld-----plsalp---qkireq 2290

Qy    173 VPDE----------------IERVL------VGIQAYLSIRK-NSSTSGLNFFHNGEFES 209
Db   2291 iedndalmddldkrqdafsavqraandviakagnkadpavrdikakleklnnlwndvqna 2350

Qy    210 SLAEKDFVEDLLDRIQSLASNGWK-LESVPRPHLSFEAQLVAGKFHELRPIKCMEQP-SP 267
Db   2351 tkkrgssliddil----svaepfwkqlnsvmktlkdleetl----------scqeppaaq 2395

Qy    268 PSDHSRAYSGKQ--KHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDV----- 320
Db   2396 pqdikkqqvalqeirheidqtkpe-veqvrrhgsnlmnmcgepdkpevkkhiedldnawd 2454

Qy    321 -LFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKA 379
Db   2455 nitalyakree------nlidamekamefhetlqnllkfltkaedkfahlgavgsdi--- 2505

Qy    380 LPGAFPAVVAGAVRALFEKISD---------LDMESRTRLILWFSHHLSNFQFIWPWEEW 430
Db   2506 ----------davkrqieqlksfkdevdphmvevealnrqavelterts--------peq 2547

Qy    431 AFVLDLP------KW-APKRVFVQEILQREVRLSYWDKIKQSI--------ENATALEEL 475
Db   2548 aasireplsvvnrrweallrgmverqkqlehallhlgqfqhalnellvwinktdstldql 2607

Qy    476 LPPKAGP--------------------NFMYSLEE-GKEKTEEQQLSAELS------RK 507
Db   2608 kpipgdpqllevelaklkvlandiqahqnsvdtlndagrqlietekgsveasttqeklrk 2667

Qy    508 VKEK-----QTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDI-----GSKSFTHL---- 553
Db   2668 lnnewkqllqkasdrqheleealreahgyiaevqdilgwlgdvdavigaskpvgglpeta 2727

Qy    554 VTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQM--------------------T 592
Db   2728 teqlerfmevyneldenrpkvetiqaqgqeyikrqnqmkvsssnlqhtlrtlkqrwdavv 2787

Qy    593 AVAIDRMMGYRLV---------SNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNR 643
Db   2788 srasdkkikleialkeatefhdtlqafvewltqae-----------kllsna--epvsr 2833

Qy    644 ISD----------LRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENP 692
```

```
Db    2834 vletiqaqmeehkvlqkdvsthreamlllдkkgth----lkyfsqkqdviliknllvsvq 2889

Qy     693 AKMKRLKS-TVEKTGEAELSLRESLEAKEAL--LNRALSETEVLLLLLFQ---------- 739
           : :|: |   |:|   :   :|: |  :|   : : | ||| :|  : :
Db    2890 hrwervvskaaertraldhgykearefndawsgmmqylqeteqvldqiieeataskepqk 2949

Qy     740 --SFLGVLKE---------RLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCE 788
             ::| |||          : | | :|: ::|     : |:|  :|        |   |
Db    2950 ikkyigklkethrqlgakqsvydgt-mrtgknllerapkgdrp-vld--------kmlie 2999

Qy     789 VGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAISSALQF 845
           : :|||          || ::  |    :  |: : ||:     |  |:   | :
Db    3000 l--keqw----------trvwsksidrqrkleealllsgq-----fsdalgelldw 3038


RESULT  10
ABB66487
ID   ABB66487 standard; Protein; 5385 AA.
XX
AC   ABB66487;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 26253.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL10590.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 26253; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   5385 AA;


  Query Match             3.1%;  Score 135;  DB 22;  Length 5385;
  Best Local Similarity   19.1%;  Pred. No. 0.8;
  Matches  183;  Conservative  156;  Mismatches  348;  Indels  270;  Gaps   43;

Qy      55 EQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCN--SIRILLRFMTSL 112
           |:|   |  :  |:  ||   |    : :  |    ||: |   ::|: ||

Db    3717 erllqsiaehkplldklnktgealgalvadddgakineildtdnaryaalrlelrerqqa 3776

Qy     113 LCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQ 172
           | | : : :      | :| : | |||:         | |     ||:||    :: ||
Db    3777 lesalqessqfsdklegmlralantvdqvn-------qld-----plsalp---qkireq 3821

Qy     173 VPDE---------------IERVL------VGIQAYLSIRK-NSSTSGLNFFHNGEFES 209
           : |                ::|       | :|  ::|   :   ||   |   :
Db    3822 iedndalmddldkrqdafsavqraandviakagnkadpavrdikakleklnnlwndvqna 3881

Qy     210 SLAEKDFVEDLLDRIQSLASNGWK-LESVPRPHLSFEAQLVAGKFHELRPIKCMEQP-SP 267
           :      ::|:|    |:|   || | || :     |  |           | | | :
Db    3882 tkkrgssllddil----svaepfwkqlnsvmktlkdleetl----------scqeppaaq 3926

Qy     268 PSDHSRAYSGKQ--KHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDV----- 320
           | |  :     |  :|:     |: : ::    :| |      |:  |:::: |:
Db    3927 pqdikkqqvalqeirheidqtkpe-veqvrrhgsnlmnmcgepdkpevkkhiedldnawd 3985

Qy     321 -LFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKA 379
           : |   |:|      ||           ||| : :  | :   |  :   |  |:
Db    3986 nitalyakree------nlidamekamefhetlqnllkfltkaedkfahlgavgsdi--- 4036

Qy     380 LPGAFPAVVAGAVRALFEKISD---------LDMESRTRLILWFSHHLSNFQFIWPWEEW 430
                  ||:   |::           :::|:  |  :  :   |           |
Db    4037 ----------davkrqieqlksfkdevdphmvevealnrqavelterts--------peq 4078

Qy     431 AFVLDLP------KW-APKRVFVQEILQREVRLSYWDKIKQSI--------ENATALEEL 475
           |  :  |      :| |  |  |:   | |  | :  : : ::        :  : |::|
Db    4079 aasireplsvvnrrweallrgmverqkqlehallhlgqfqhalnellvwinktdstldql 4138

Qy     476 LPPKAGP--------------------NFMYSLEE-GKEKTEEQQLSAELS------RK 507
           |   |                    | : :| : |::  | :: | | |       ||
Db    4139 kpipgdpqllevelaklkvlandiqahqnsvdtlndagrqlietekgsveasttqeklrk 4198

Qy     508 VKEK-----QTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDI-----GSKSFTHL---- 553
           :  :     | | |     :|| :   ||:   :  ::  | |:      ||    |
Db    4199 lnnewkqllqkasdrqheleealreahgyiaevqdilgwlgdvdavigaskpvgglpeta 4258

Qy     554 VTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQM--------------------T 592
              |||: :|:::|  :  |   : :|   | |   ||
Db    4259 teqlerfmevyneldenrpkvetiqaqgqeyikrqnqmkvsssnlqhtirtlkqrwdavv 4318

Qy     593 AVAIDRMMGYRLV---------SNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNR 643
           : | |: :   :          : || | |:   |             ::| ||  :  :|
Db    4319 srasdkkikleialkeatefhdtlqafvewltqae-----------kllsna--epvsr 4364

Qy     644 ISD-----------LRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENP 692
           : :           |:||:|   : :|: :|   :    |:    |  ::  : :|
Db    4365 vletiqaqmeehkvlqkdvsthreamlllдkkgth----lkyfsqkqdviliknllvsvq 4420

Qy     693 AKMKRLKS-TVEKTGEAELSLRESLEAKEAL--LNRALSETEVLLLLLFQ---------- 739
           : :|: |   |:|   :   :|: |  :|   : : | ||| :|  : :
Db    4421 hrwervvskaaertraldhgykearefndawsgmmqylqeteqvldqiieeataskepqk 4480

Qy     740 --SFLGVLKE---------RLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCE 788
             ::| |||          : | | :|: ::|     : |:|  :|        |   |
Db    4481 ikkyigklkethrqlgakqsvydgt-mrtgknllerapkgdrp-vld--------kmlie 4530

Qy     789 VGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAISSALQF 845
           : :|||          || ::  |    :  |: : ||:     |  |:   | :
Db    4531 l--keqw----------trvwsksidrqrkleealllsgq-----fsdalgelldw 4569


RESULT  11
ABB67161
ID   ABB67161 standard; Protein; 5496 AA.
XX
AC   ABB67161;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 28275.
XX
```

```
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL11264.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 28275; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   5496 AA;
```

```
  Query Match             3.1%;  Score 135;  DB 22;  Length 5496;
  Best Local Similarity   19.1%;  Pred. No. 0.82;
  Matches  183;  Conservative  156;  Mismatches  348;  Indels  270;  Gaps   43;

Qy      55 EQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCN--SIRILLRFMTSL 112
           |:|   |  :  |:  ||   |     : :   |     ||: |    ::|: ||
Db    3828 erllqsiaehkplldklnktgealgalvadddgakineildtdnaryaalrlelrerqqa 3887

Qy     113 LCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQ 172
           | | : : :      | :| : | |||:         | |      ||:||    :: ||
Db    3888 lesalqessqfsdklegmlralantvdqvn-------qld-----plsalp---qkireq 3932

Qy     173 VPDE---------------IERVL------VGIQAYLSIRK-NSSTSGLNFFHNGEFES 209
           : |                 ::|         | :|  ::|   :    ||   |    :
Db    3933 iedndalmddldkrqdafsavqraandviakagnkadpavrdikakleklnnlwndvqna 3992

Qy     210 SLAEKDFVEDLLDRIQSLASNGWK-LESVPRPHLSFEAQLVAGKFHELRPIKCMEQP-SP 267
           :      ::|:|    |:|   || | || :     |  |                | | : 
Db    3993 tkkrgssliddil----svaepfwkqlnsvmktlkdleetl----------scqeppaaq 4037

Qy     268 PSDHSRAYSGKQ--KHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDV----- 320
           | |  :     |  :|:     |: : ::    :| |      |:  |:::: |:
Db    4038 pqdikkqqvalqeirheidqtkpe-veqvrrhgsnlmnmcgepdkpevkkhiedldnawd 4096

Qy     321 -LFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKA 379
            :  |   |:|      ||          ||| : :  | :   |  :   |  |:
Db    4097 nitalyakree------nlidamekamefhetlqnllkfltkaedkfahlgavgsdi--- 4147

Qy     380 LPGAFPAVVAGAVRALFEKISD---------LDMESRTRLILWFSHHLSNFQFIWPWEEW 430
                  ||:   |::          :::|:  |  :  :   |         |
Db    4148 ----------davkrqieqlksfkdevdphmvevealnrqavelterts--------peq 4189
```

```
Qy     431 AFVLDLP------KW-APKRVFVQEILQREVRLSYWDKIKQSI--------ENATALEEL 475
           |  :  |      :| |  |  |:   | |  | :  : : ::        :  : |::|
Db    4190 aasireplsvvnrrweallrgmverqkqlehallhlgqfqhalnellvwinktdstldql 4249

Qy     476 LPPKAGP--------------------NFMYSLEE-GKEKTEEQQLSAELS------RK 507
             |    |                    | : :| : |::  | :: | | |       ||
Db    4250 kpipgdpqllevelaklkvlandiqahqnsvdtlndagrqlietekgsveasttqeklrk 4309

Qy     508 VKEK-----QTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDI-----GSKSFTHL---- 553
           :  :     | | |    :|| :   ||:   :  ::  | |:      ||    |
Db    4310 lnnewkqllqkasdrqheleealreahgyiaevqdilgwlgdvdavigaskpvgglpeta 4369

Qy     554 VTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQM--------------------T 592
             |||: :|:::|  :  |   : :|   | |   ||
Db    4370 teqlerfmevyneldenrpkvetiqaqgqeyikrqnqmkvsssnlqhtlrtlkqrwdavv 4429

Qy     593 AVAIDRMMGYRLV---------SNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNR 643
           : | |: :   :           : || | |:   |            ::| ||  :  :|
Db    4430 srasdkkikleialkeatefhdtlqafvewltqae-----------kllsna--epvsr 4475

Qy     644 ISD----------LRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENP 692
           : :           |:||:|   : :|: :|   :    |:    |  ::  : :|
Db    4476 vletiqaqmeehkvlqkdvsthreamllldkkgth----lkyfsqkqdvilikn|lvsvq 4531

Qy     693 AKMKRLKS-TVEKTGEAELSLRESLEAKEAL--LNRALSETEVLLLLLFQ---------- 739
           : :|: |   |:|   :   :|: |  :|   : : | ||| :|  : :
Db    4532 hrwervvskaaertraldhgykearefndawsgmmqylqeteqvldqiieeataskepqk 4591

Qy     740 --SFLGVLKE---------RLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCE 788
             ::| |||          : | | :|: ::|     : |:|  :|        |   |
Db    4592 ikkyigklkethrqlgakqsvydgt-mrtgknllerapkgdrp-vld--------kmlie 4641

Qy     789 VGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAISSALQF 845
           :  :|||           || ::  |    :  |: : ||:     |  |:   | :
Db    4642 l--keqw----------trvwsksidrqrkleealllsgq----fsdalgelldw 4680
```

```
RESULT  12
ABB67112
ID   ABB67112 standard; Protein; 8805 AA.
XX
AC   ABB67112;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 28128.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL11215.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
```

```
PS   Disclosure; SEQ ID NO 28128; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   8805 AA;


  Query Match           3.1%;  Score 135;  DB 22;  Length 8805;
  Best Local Similarity  19.1%;  Pred. No. 1.8;
  Matches  183;  Conservative  156;  Mismatches  348;  Indels  270;  Gaps   43;

Qy      55 EQLPHKIPLYGTLIGLLNLENEDFVQKLVESVHANFQVALDSGNCN--SIRILLRFMTSL 112
           |:|   |  :  |:  ||   |     : :   |     ||: |    ::|: ||
Db    7132 erllqsiaehkplldklnktgealgalvadddgakineildtdnaryaalrlelrerqqa 7191

Qy     113 LCSKVIQPASLIVVFETLLSSAATTVDEEKGNPSWQPQADFYVICILSSLPWGGSELAEQ 172
           | | : : :      | :| : | |||:         | |      ||:||    :: ||
Db    7192 lesalqessqfsdklegmlralantvdqvn------qld-----plsalp---qkireq 7236

Qy     173 VPDE---------------IERVL------VGIQAYLSIRK-NSSTSGLNFFHNGEFES 209
           : |                 ::|         | :|  ::|   :    ||   |    :
Db    7237 iedndalmddldkrqdafsavqraandviakagnkadpavrdikakleklnnlwndvqna 7296

Qy     210 SLAEKDFVEDLLDRIQSLASNGWK-LESVPRPHLSFEAQLVAGKFHELRPIKCMEQP-SP 267
           :      ::|:|    |:|   || | || :     |  |           | | | :
Db    7297 tkkrgsslddil----svaepfwkqlnsvmktlkdleetl----------scqeppaaq 7341

Qy     268 PSDHSRAYSGKQ--KHDALTRYPQRIRRLNIFPANKMEDVQPIDRFVVEEYLLDV----- 320
           | |  :     |  :|:     |: : ::    :| |      |:  |:::: |:
Db    7342 pqdikkqqvalqeirheidqtkpe-veqvrrhgsnlmnmcgepdkpevkkhiedldnawd 7400

Qy     321 -LFYLNGCRKECASYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYYTLVIMDLCKA 379
            :  |   |:|      ||          |||  :  :  |  :   |  :   |  |:
Db    7401 nitalyakree------nlidamekamefhetlqnllkfltkaedkfahlgavgsdi--- 7451

Qy     380 LPGAFPAVVAGAVRALFEKISD---------LDMESRTRLILWFSHHLSNFQFIWPWEEW 430
                  ||:   |::          :::|:  |  :  :   |        |
Db    7452 ----------davkrqieqlksfkdevdphmvevealnrqavelterts--------peq 7493

Qy     431 AFVLDLP------KW-APKRVFVQEILQREVRLSYWDKIKQSI--------ENATALEEL 475
           |  :  |      :| |  |  |:   | |  | :  : : ::        :  : |::|
Db    7494 aasireplsvvnrrweallrgmverqkqlehallhlgqfqhalnellvwinktdstldql 7553

Qy     476 LPPKAGP--------------------NFMYSLEE-GKEKTEEQQLSAELS------RK 507
            |    |                    | : :| : |::  | :: | | |      ||
Db    7554 kpipgdpqllevelaklkvlandiqahqnsvdtlndagrqlietekgsveasttqeklrk 7613

Qy     508 VKEK-----QTARDMIVWIEETIYPVHGFEVTLTIVVQTLLDI-----GSKSFTHL---- 553
           :  :      | | |    :|| :   ||:   :  ::  | |:      ||    |
Db    7614 lnnewkqllqkasdrqheleealreahgyiaevqdilgwlgdvdavigaskpvgglpeta 7673

Qy     554 VTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQM--------------------T 592
              |||: :|:::|  :  |   : :|   | |   ||
Db    7674 teqlerfmevyneldenrpkvetiqaqgqeyikrqnqmkvsssnlqhtlrtlkqrwdavv 7733

Qy     593 AVAIDRMMGYRLV---------SNQAIVRWVFSPENVDQFHVSDQPWEILGNALNKTYNR 643
           : | |: :   : ·         : || | |:   |            ::| ||  :  :|
Db    7734 srasdkkikleialkeatefhdtlqafvewltqae-----------kllsna--epvsr 7779

Qy     644 ISD----------LRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENP 692
           : :           |:||:|   : :|: :|   :    |:   |  ::  : :|
Db    7780 vletiqaqmeehkvlqkdvsthreamlldkkgth----lkyfsqkqdviliknllvsvq 7835

Qy     693 AKMKRLKS-TVEKTGEAELSLRESLEAKEAL--LNRALSETEVLLLLLFQ---------- 739
           : :|: |   |:|   :   :|: |  :|   : : | ||| :|  : :
Db    7836 hrwervvskaaertraldhgykearefndawsgmmqylqeteqvldqiieeataskepqk 7895

Qy     740 --SFLGVLKE---------RLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSCE 788
             ::| |||           : | | :|: ::|     : |:|   :|        |   |
Db    7896 ikkyigklkethrqlgakqsvydgt-mrtgknllerapkgdrp-vld--------kmlie 7945

Qy     789 VGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAISSALQF 845
           :  :|||           || ::  |    :  |: : ||:     |  |:   | :
Db    7946 l--keqw----------trvwsksidrqrkleealllsgq-----fsdalgelldw 7984


RESULT  13
AAB93096
ID   AAB93096 standard; Protein; 1084 AA.
XX
AC   AAB93096;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:11947.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 11947; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
```

```
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   1084 AA;


  Query Match             3.1%;  Score 134.5;  DB 22;  Length 1084;
  Best Local Similarity   17.9%;  Pred. No. 0.061;
  Matches  162;  Conservative  143;  Mismatches  307;  Indels  293;  Gaps   37;

Qy      34 VIRREIERSGDQVLPFLLQCAEQLPHKIPLYGTLI--------------GLLNLENEDF 78
           |: | : |:   :  ||      |   | :|  |:              | |:
Db     265 vlnrvvtrasspishfl-----dllsnivmyaplvlqscsskvteafdylsflplqt--- 316

Qy      79 VQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLIVVFETLLSSAATTV 138
           ||:|:::|    :|::   :|    ::|    ::  :::    | :  |  ||
Db     317 vqrllkavqpllkvsmsmrdc----lilvlrkamfanqldarksavagfllllk------ 366

Qy     139 DEEKGNPSWQPQADFYVICILSSLPWGGSELAEQVPDEIERVLVGIQAYLSIRKNSSTSG 198
                     :| |:  |||     |: :: :   : :| | : ::
Db     367 ------------nfkvlgslss-----sqcsqsl--svsqvhvdvhshy---------- 396

Qy     199 LNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLVAGKFHELR- 257
                  :|:| : |  :::| ::   |                  |  | :  ||
Db     397 ----------nsvanetfcleimdslrrclsq----------qadvrltlyegfydvlrr 436

Qy     258 --------------PIKCMEQPSPPSDHSRAYSGKQKHDALTRYPQRIRRLNIFPANKME 303
                         :|   :| |              | |   | ::    :   :|:
Db     437 nsqlansvmqtllsqlkqfyepkp--------------dllp--plkleaciltqgdkis 480

Qy     304 DVQPIDRFVVEEYLLDVLFYLNGCRKECASYMANLPVTFR------------YEYL--MA 349
             :|:|              ||  | : | ::  |  :  :             || |  :
Db     481 lqepld------------yllcciqhclawyknt viplqqgeeeeeeeeafyedlddil 527

Qy     350 ETLFSQILLLPQPPF----------------KTLYYTLVIMDLCKAL-----------PG 382
           |:: ::::      |        .        |      ::| :|: |
Db     528 esitnrmikseledfeldksadfsqstsigiknnisaflvmgvcevlieynfsissfskn 587

Qy     383 AFPAVVAGAVRALFEKISDLDME----SRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPK 438
            |  ::  ::   ::|:||:  |    ::|::    |  | : :
Db     588 rfedil--slfmcykklsdilnekagkaktkmanktsdsllsmk---------------- 629

Qy     439 WAPKRVFVQEILQREVRLSYWDKIKQSIENATALEELLPPKAGPNFMYSLEEGKEK---- 494
                ||  :|    |:  |    :     |::||    : | :      || :|
Db     630 ------fvssllталfrvllw----rytsiptsveesgkkekgksisllcleglqkifsa 679

Qy     495 ------TEEQQLSAELSRKVKEKQTARDMIVWI-EETIYPVHGFEVT-LTIVVQTLLDIG 546
                : ||    |    || :   |  | : : | : :  |: : | ::     |
Db     680 vqqfyqpkiqqflraldvtdkegeeredadvsvtqrtafqirqfqrsllnllssqeedfn 739

Qy     547 SKSFTHLVTVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVS 606
           ||    ||||    ::   | | : : | :||  |   | | :  |:    :|
Db     740 skealllvtvltslskl---lepsspqfvqmlswtskickensredalfckslmnl---- 792

Qy     607 NQAIVRWVFSPENVDQFHVS-DQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKAS 665
                :||       |||   |  :| :     :  : |: :|:              || :
Db     793 -------lfs------lhvsykspvillrdlsqdihghlgdidqdve--------vektn 831

Qy     666 ANARVELEAAESKLSLV---EGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEAL 722
           | | |  |   : |:   : | || |    : :||  | :   :| |      ::  |
Db     832 hfaivnlrtaaptvcllvlsqaekvleevdwlitklkgqvsqetlseas------sqatl 885

Qy     723 LNRALSETEVLLL-LLFQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENG 781
           |: : :  :: |  |   |  :::  ||  : | ::  || :
Db     886 pnqpvekaiimqlgtlltffhelvqtalpsgscvdtl--lkdl---------------- 926

Qy     782 NPKKSCEVGEREQWCLSTLGYLTAFTRQY------ASEIWPHMEKLESEVFSGEDVHPLF 835
                |::           |||  | |      :  | :||||     ||  : ||
Db     927 -----ckm----------yttltalvryylqvcqssggipknmeklvk--lsgshltplc 969

Qy     836 LQAIS 840
             ||
Db     970 ysfis 974
```

```
RESULT  14
AAM78497
ID   AAM78497 standard; Protein; 710 AA.
XX
AC   AAM78497;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 1159.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK51630.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 20; Page 3397-3398; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence   710 AA;


  Query Match             3.1%;  Score 133.5;  DB 22;  Length 710;
  Best Local Similarity   21.0%;  Pred. No. 0.037;
  Matches   86;  Conservative   74;  Mismatches  163;  Indels   87;  Gaps   16;

Qy     461 KIKQSIENAT--------------ALEELLPPKAGPNFMYSLEEGKEKTEEQQLSAELS 505
           |:|| :|:|:                |:| :           :  :|:::   :  ||
Db     127 kmkqdledasnkaeeerarlegelkglqeqiaetkarlitqqhdraqeqsdhalmlrelq 186
```

```
Qy     506 RKVKEKQTAR-DMIVWIEET----------IYPVHGFEVTLTIVVQTLLDIGSKSFTHLV 554
Db     187 kllqeertqrqdlelrleetrealagrayaaeqmegfelqtkqltreveelkse----lq 242

Qy     555 TVLERYGQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAIDRMMGYRLVSNQAIVRWV 614
Db     243 airdeknqpdprlqelqeeaarlkshfqaqlqqemrktalaedqlrqqsqveeqrva--- 299

Qy     615 FSPENVDQFHVSDQPWEILGN----------ALNKTYNRISDLRKDISNITKNVLVAEKA 664
Db     300 -alen----qisevs-ellgtyekakqkdqlaiqklkeri--lqldlenkt----lalaa 347

Qy     665 SANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEAELSLRESLEAKEALLN 724
Db     348 ssrspldshgeessldv----nvlkdkmeklkrllqvaarksqvtldv-eklcdleimps 402

Qy     725 RALSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPK 784
Db     403 seaadgekatalyyqqelkqlkeef-erykmraqvvlkskntkd-----------gnlg 449

Qy     785 KSCEVGEREQWCLSTLGYLTAFTRQYAS-----EIWPHMEKLESEVFSGE 829
Db     450 keleaaqeq---------laelkekyislrlsceelehqhqqeaddwkqe 490
```

RESULT 15
ABB67210
ID ABB67210 standard; Protein; 3263 AA.
XX
AC ABB67210;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 28422.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL11313.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 28422; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3263 AA;

Query Match 3.1%; Score 133.5; DB 22; Length 3263;
Best Local Similarity 18.7%; Pred. No. 0.47;
Matches 180; Conservative 142; Mismatches 321; Indels 321; Gaps 48;

```
Qy      74 ENEDFVQKLVESVHANFQVALDSGNCNSIRILLRFMTSLLCSKVIQPASLIVVF---ETL 130
Db    2175 esprfveelvqpvevmdgeal---------------lltcqvtgkptpkvewyhnaeki 2218

Qy     131 LSSAATTVDEE-KGNPSWQ-----PQADFYVICILSSLPWGGSELAEQVPDEIERVLVGI 184
Db    2219 tenkettisqdlqgvcqlqitevfpenegqyecvatn----------kigksvsktnvki 2268

Qy     185 QAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIQSLASNGWK-LESVP---RP 240
Db    2269 qafeyi-pdseitglt-------------gseedlldrtlsideqapkiikklpekiep 2313

Qy     241 HLSFEAQLVAGKFHELRPIKCMEQPSPP----SDHSRAYSGKQKHDALTRYPQRIRRLNI 296
Db    2314 kegeqakl---------evkvvgkpkpkvkwlrddeqifasee---------------- 2347

Qy     297 FPANKMED---------VQPIDRFVV--EEY------LLDVLFYLNGC-------RKECA 332
Db    2348 yqienfedgtsvlvinhvypddlgtisfeaynplgvavttalfavegivgskdyrkpewv 2407

Qy     333 SYMANLPVTFRYEYLMAETLFSQILLLPQPPFKTLYY----TLVIMDLCKALPGAFPAVV 388
Db    2408 sqmeemqvalk----diqwyfnnvqlrasekyrmvvqeqeatleimkitsedcgyynckl 2463

Qy     389 AGAV-----RALFEKISDLDMESRTRLILWFSHHLSNFQFIWPWEEWAFVLDLPKWAPKR 443
Db    2464 ineigmtmtrakf------disststiv----------------eetkakttvkkksgkk 2501

Qy     444 VFV--------QEILQREVRLSYWD---------KIKQSIE---NATALEELLPPKAGPN 483
Db    2502 tmvkrsgasesqnvqkteiriiptsavetsmnvikvkqpvsvlvekseisevlvvk---- 2557

Qy     484 FMYSLEEGKEKTEEQQLSAELSRKVKEKQTARDMI---------VWIE--ETIYPVHGFE 532
Db    2558 -------drevadaeerssqlieeieeeeeieekvqhdeedevevqveqketytsskkie 2610

Qy     533 VTLTI----------------------VVQTLLD---------IGSKSFTHLVTV--L 557
Db    2611 itktvelirtkisekiitiedvqvlshheevqwllesieaesfgqigesalrdlatigll 2670

Qy     558 ERYG----------QVFSKL-CPDNDKQVM---------LLSQVSTYWKNNVQMTAV 594
Db    2671 lrygcehyeitymyeqnifislkkpesqsalvqlveregheelisqilsessnedetila 2730

Qy     595 AIDRMMGY----RLVSNQAIVRWVFSPENVDQFHVSDQPWEILGN---------ALNKT 640
Db    2731 a---tvgfkpfirmiqtyeitieivirkfvredfis-qdwkicgkeeeeilpqkvlainee 2786

Qy     641 YNRISD--LRKDI-----SNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPA 693
Db    2787 vlpldefglrkesprpkenkltenievrlkhalnvs-haktaesskelpskipksvkaqr 2845

Qy     694 KMKRLKSTVEKTGEAELSL-------------------------RESLEAKEALLNRA 726
Db    2846 kmkesrslvveapnaeeaiedlkplkavsqevqsdilfsheiteeqhqaletieklkpts 2905

Qy     727 LSETEVLLLLLFQSFLGVLKERLPDPTKVRSVQDLKSIGAEDDKPSA---MDVDSENGNP 783
Db    2906 aiedtvqqkllsqeel-iiaevlpsetvgrdvtdvrppg-etisprltpnmslcitecqp 2963

Qy     784 KKSCEVGEREQWCLSTLGYLTAFTRQYASEIWPHMEKLESEVFSGEDVHPLFLQAISSAL 843
Db    2964 eds--igemqq---------aakermet---psmsvteskavggqe-----levlenvd 3003
```

```
Qy     844 QFPL 847
             ||
Db    3004 hmpl 3007
```

Search completed: August 20, 2002, 15:25:52
Job time: 5641 sec